

# Simulated Real-Time Detection of a Small Molecule on a Carbon Nanotube Cantilever

Reese E. Jones<sup>1,\*</sup>, Jeremy A. Templeton<sup>1</sup>, and Timothy W. Rebold<sup>2</sup>

<sup>1</sup>*Sandia National Laboratories, Livermore, CA 94551, USA*

<sup>2</sup>*Purdue University, West Lafayette, IN 47907, USA*

In this work we explore limits of using a functionalized carbon nanotube as a mechanical sensor of small gas-phase molecules. Specifically, we investigate how fast can a change in tip mass be detected and how small a change can be detected as a simplified version of the attachment of weakly-bonded gas phase molecules. Filtering and identification techniques are based on a specialized, parallel Kalman filter, newly developed continuum beam theory and a Langevin model of the thermal vibrations of the carbon nanotube. In this simulation-based work, we account for thermal noise that is intrinsic to the carbon nanotube and the attached substrate but omit detector noise for simplicity.

**Keywords:** Sensor, Carbon Nanotube, Kalman Filter, Beam Theory, Langevin Model.

Sun, 21 Aug 2011 09:41:12

## 1. INTRODUCTION

Over the last few decades, nanomechanical systems have drawn keen interest and have started to make technological impact (see Refs. [1, 2] for a review). Specifically, sensors have become more accurate and sensitive as their primary components have been scaled down. For example, cantilever-based devices, like the atomic force microscope, have benefited from less inertia as their tips have been made smaller. These devices have also experienced increased sensitivity to force due to several factors peculiar to the nanoscale, such as reduced inertia.

A mass sensor based on a carbon nanotube (CNT) has the potential to be extraordinarily sensitive, as the recent work of Jensen, Kim and Zettl<sup>3</sup> and others<sup>4–7</sup> demonstrates. As evidenced by the work of Treacy et al.<sup>8</sup> and others, e.g., Ref. [9]; however, these nano-scale devices are subject to substantial thermally-driven vibrations. This is particularly problematic since most of these mechanical sensors depend on their resonance properties<sup>9–14</sup> to function. As a result, many of the experiments were performed at cryogenic temperatures<sup>15,16</sup> to reduce thermal noise or forced using an external electric field<sup>17</sup> to enhance the signal relative to the thermal noise. Vibrations due to thermal energy are not the only source of noise in these cantilever sensors but it is one of the few that is intrinsic to the CNT. A number of authors<sup>2,18,19</sup>

have explored the various sources of noise and associated dissipation mechanisms in a comprehensive fashion. In addition to design of the nanoscale cantilever and the functionalization required to capture the target molecule, the construction of the detector that identifies changes the resonator's motion involves considerable technical development. There have been many complex approaches to tip displacement detectors, e.g., field emission,<sup>3,17</sup> electron microscopy, laser interferometry,<sup>15</sup> piezoelectric changes, changes in capacitance,<sup>12</sup> and techniques based on a single electron transistor.<sup>16,20,21</sup> The optimal technique depends on the material, environment and the configuration, and continues to be the subject of considerable research. If these challenges can be overcome, CNT-based sensors could enable a revolution in sensing technology in the scientific, industrial and public safety arenas. For instance, on-chip arrays of sensors could be used to detect chemicals in low concentrations and on short timescales. Moreover, they could be deployed in a widely dispersed and locally redundant fashion with different functionalization specific to multiple molecules.

The focus of this work is to examine the limits of real-time detection in the presence of thermal noise; hence we assume a perfect, noiseless detector in order to concentrate on the intrinsic thermal noise of the CNT-substrate system. To fully realize the potential of a CNT-based sensor in a gaseous environment we develop real-time detection techniques for molecules of interest on a functionalized tip within a finite residence time. Our intent is to improve

\*Author to whom correspondence should be addressed.

practical detection algorithms such as that of Jensen et al.<sup>3</sup> which is based on the simple (sensitivity) relation

$$\Delta\omega \approx \left. \frac{\partial\omega}{\partial m} \right|_{m=0} m = -2 \frac{m}{M} \omega$$

between the observed frequency  $\omega$  and the ratio of the added mass  $m$  to the mass of the CNT  $M$ .<sup>a</sup> The intrinsic sensitivity  $\partial\omega/\partial m$  is commonly referred to as “mass responsivity,” see, e.g., Ref. [14]. In Jensen et al.’s pioneering work, the authors show that they can detect from 1 to 51 Au atoms adhered to a multi-wall CNT of length 205 nm and mass  $M = 1.58 \times 10^{-21}$  kg, corresponding to mass ratios of  $m/M \approx 0.00021$  for 1 attached Au atom, and  $m/M \approx 0.0011$  for 51 attached Au atoms. These measurements are based on a technique detailed in Ref. [5] where the spectral density of the observed, fluctuating frequency is integrated over a frequency bin around the fundamental frequency in order to detect a frequency shift in the resonator. Fundamentally, mass resolution depends on observation time since the width of the bin is inversely proportional to this parameter. We estimate the detection time associated with the Jensen et al. device<sup>3</sup> to be on the order of 1 second. Time-resolution on the order of 1 second is commensurate with the work of Feng et al.,<sup>14</sup> which also observed that the mass resolution behaves like a power law with observation time for their Si nanowire device.

Using the computational technique of molecular dynamics (MD) we have the capacity to simulate the thermal vibrations of a CNT and the effects of an attached molecule with atomic detail and on time-scales ranging from the period of atomic vibrations to nanoseconds. Specifically, we employ an MD model system of the CNT embedded in a substrate with or without a molecule attached to the tip, see Figure 1. Due to (hypothetical) selective functionalization of the tip only, we assume that the binding of a small molecule occurs only at the tip. An explicit, thermalized substrate is chosen so that the CNT will have trajectories that have realistic thermal fluctuations on time-scales ranging from that of atomic vibrations to the fundamental period of the cantilevered CNT without the direct application of a thermostat or ensemble averaging.

To facilitate this approach, many simplifications are expedient and necessary to make the problem computationally tractable. Foremost, is how the molecule is and becomes attached to the tip of the CNT. In this work, we add a representative mass directly by increasing the

mass of the atoms at the free end of the cantilevered CNT. Since the resonant frequencies of a small attached molecule would be very high compared to the fundamental frequency of beam,<sup>b</sup> we believe this assumption does not change the fundamental nature of the problem. Also, in this work we do not simulate the complex dynamics of capture of a molecule by a functionalized tip. Specifically, we do not attempt to model mechanics of collision (energy and momentum transfer) and attachment (the energetics of bonding) nor the exact mechanisms resulting in a finite residence time in the gas phase (see Ref. [6] for a discussion). We also omit fluid damping on the vibrations of the CNT since we assume it is negligible at this scale (see Ref. [13] for experimental data of a CNT resonator in air and vacuum where the differences are ascribed to adsorbed molecules and the adsorption-desorption noise described in Ref. [22]).

With these assumptions and limitations, the problem we examine is reduced to: given a CNT of a well-characterized length, radius, and flexural rigidity at the ambient temperature (a) what is the minimum mass that can be detected for a given error tolerance, (b) what is the minimum time for detection given a mass. Our proposed solution is to use a parallel array of Kalman filters (KFs) deployed as an “on-the-fly” Fourier transform in the neighborhood of the fundamental resonance frequency. We compare the results of this adaptive KF algorithm to the “truth” of an exhaustive discrete time Fourier transform (DTFT) which involves a fine-grained frequency scan over the range of interest.

To connect our solution to the problem at hand, we first develop the theory to describe the shift in amplitude and frequency based on an attached mass in Section 2. In this section, we provide a coarse-grained continuum beam model that provides an understanding and prediction of the resonance structure of a cantilevered CNT. Building on this, we present a generalization of the results of Treacy et al. to a CNT with an added mass. This beam-based theory gives us an accurate model of the shifts in the resonant frequency and amplitude in response to an attached molecule. In addition, we develop a reduced-order Langevin model of the CNT vibrations to describe the shape and intensity of the primary resonance peak, as well as provide a reasonable process surrogate for testing of our algorithms. Together these theories provide the process characterization necessary to relate measurable changes to added mass. Since these changes are fundamentally spectral in nature and yet we want rapid, real-time detection of an attached molecule, in Section 3 we choose to develop a specialized version of the traditional KF, as well as an adaptive algorithm to track the shift in peak amplitude due to an attached mass. In this section, we also contrast

<sup>a</sup>In these works, the CNT and added mass system is modeled as a harmonic oscillator  $\omega = \sqrt{\tilde{K}/(M + \zeta m)}$ , where  $\tilde{K}$  is an effective spring constant. The mass responsivity  $\Delta\omega$  results from a linearization of this relation. The weight factor  $\zeta$  is the ratio of the square displacement at the point of attachment to the average square displacement and  $\zeta = 4$  for a mass attached to the tip, as will be shown in Section 2.

<sup>b</sup>In addition, these frequencies are most likely at or beyond the limits of what is possible with the currently-available, appropriate frequency detection technology.

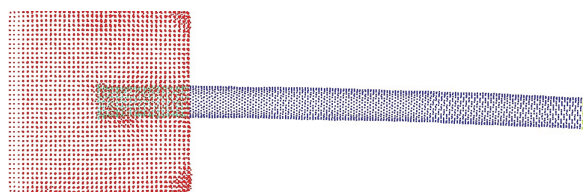
the new method with filters based on recursive Fourier transforms and sequential Bayesian updates. In Section 4 we apply the easily parallelizable algorithm to a sequence of resonator models of increasing complexity and fidelity, starting with a noisy sine, through the Langevin model, and culminating with the MD data. In this Results section we quantify the performance of the algorithm with respect to noise tolerance, sampling efficiency, etc. and answer our two thesis questions on the theoretical limits of mass detection. In the final Discussion section we summarize the main results of the paper and give indications for future work.

## 2. THERMAL VIBRATIONS OF A CARBON NANOTUBE

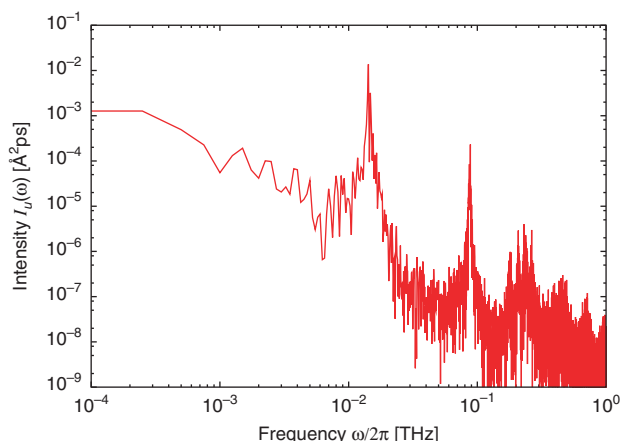
The thermal vibrations of a carbon nanotube suspended by one end are clearly visible in microphotographs, e.g., Ref. [8,9,16], and can be large relative to the diameter of the CNT at room temperature depending on its length. These fluctuations are simply the thermal lattice vibrations, i.e., phonon modes, that have amplitudes that are significant on the scale of the diameter of the CNT. It is these motions which, in part, limit the sensitivity of nanoscale sensors. The mitigation strategy proposed in this work is based on characterizing the phonon spectra followed by devising filtering techniques to enhance the signal generated by an attached mass. Fully characterizing CNT phonon spectra is a difficult endeavor, so we will instead use a hierarchy of models to provide sufficient insight into CNT thermal motions for filtering purposes.

Molecular dynamics (MD) is typically employed to model the dynamics of molecular systems on this scale and is an ideal tool for representing phonons and thermal noise, and therefore is our highest fidelity model, albeit a computationally expensive one, of the finite temperature CNT. Figure 1 shows the cantilevered configuration of the CNT/substrate/tip-mass system. Figure 2 shows the spectral response of the freely vibrating CNT tip which is key to the filtering problem we will undertake in Section 3. More details relevant to the MD version of the CNT will be given in the Results section.

However, the expense associated with MD makes it impractical for real-time detector applications. Therefore, in this section we develop two reduced order models of the thermal vibrations of a CNT: one based on continuum



**Fig. 1.** A 17.8 nm long, (10, 10) CNT (blue) embedded in a substrate of Au atoms (red). The tip atoms are highlighted in green.



**Fig. 2.** The tip displacement intensity spectrum for a 17.8 nm long, (10, 10) CNT showing a fundamental resonance at 0.0143 THz and higher resonances of decreasing intensities.

beam theory and the other on the Langevin formalism. These two models capture the salient features of the resonance of a thermalized CNT with a tip mass. The idea of using beam theory<sup>8,9,23</sup> or a full continuum model<sup>24,25</sup> to represent the motion of a CNT is well established, at least in the context of mechanical deformation (ignoring thermal vibrations). The use of a Langevin model, on the other hand, is not seen in the relevant literature for CNTs other than in its implicit use in modeling the quality factor of resonators<sup>26</sup> but does have a long history in the anharmonic phonon interaction literature.<sup>27–29</sup>

### 2.1. Beam Theory

The simplest continuum beam theory, the Euler–Bernoulli beam

$$\rho A \ddot{\mathbf{u}} = - \frac{\partial^2}{\partial x^2} \left( EI \frac{\partial^2}{\partial x^2} \mathbf{u} \right) \quad (1)$$

is appropriate for relatively long, narrow beams and small (linear), low frequency vibrations. It is sufficient to describe the first/fundamental mode of CNTs.<sup>30</sup> We denote  $\mathbf{u}(x, t)$  as the transverse deflection with  $x$  being the coordinate along the length of the beam,  $\mathbf{u}' = (\partial/\partial x)\mathbf{u}$  is the slope of the deflection,  $EI\mathbf{u}''$  is the bending moment, and  $-(EI\mathbf{u}'')$  is the shear force. Here,  $\rho A$  is linear mass density and  $EI$  is the flexural rigidity.<sup>c</sup> The constitutive model relates the local curvature to the bending moment and ignores the energy from all other deformation modes. The dispersion relation between frequency  $\omega$  and wave-number  $k$  is

$$\omega^2 = \frac{EI}{\rho A} k^4 \quad (2)$$

<sup>c</sup>  $E$  is Young's modulus and  $I$  is the area moment of inertia. For a tube  $I = (\pi/4)(r_o^4 - r_i^4) = (\pi/4)A(r_o^2 + r_i^2)$ . Since  $I$  is ambiguous for a CNT due to the need to define a thickness  $r_o - r_i$ , we consider only the flexural rigidity  $EI$  which is a constant for a given chirality.

For the case of a point mass affixed to the end of a Euler–Bernoulli cantilever, the shear boundary condition at  $x = L$  is

$$\frac{\partial}{\partial x} \left( EI \frac{\partial^2 \mathbf{u}}{\partial x^2} \right) \Big|_{x=L} = \mathbf{f} \quad (3)$$

where  $\mathbf{f}$  is the force a concentrated mass exerts on the tip. The other boundary conditions are  $\mathbf{u} = \mathbf{0}$  and  $\mathbf{u}' = \mathbf{0}$  at  $x = 0$  and  $-EI\mathbf{u}'' = \mathbf{0}$  at  $x = L$ . Since the attached mass at the tip can be treated simply as a boundary condition and the governing equation has not changed, the assumed modal solution for this fixed-free beam is the same as it would be without the attached mass:

$$\mathbf{u}(x, t) = \sum_{n=0}^{\infty} \mathbf{u}_n(t) \chi_n(x) \quad (4)$$

where  $\chi_n = \chi(k_n x) = A_1 \cos(k_n x) + A_2 \sin(k_n x) + A_3 \cosh(k_n x) + A_4 \sinh(k_n x)$ . The mode shapes  $\chi_n$  that satisfy the two boundary conditions at  $x = 0$  are

$$\chi_n = A_1 (\cos(k_n x) - \cosh(k_n x)) + A_2 (\sin(k_n x) - \sinh(k_n x)) \quad (5)$$

In the case of the mass being directly attached to the beam,<sup>d</sup> Newton's law (3) for the small molecule reduces to

$$\mathbf{f} = m\ddot{\mathbf{u}}|_{x=L} = m\omega^2 \mathbf{u}|_{x=L} = EIL \frac{m}{M} k^4 \mathbf{u} \Big|_{x=L} \quad (6)$$

after substitution of the dispersion relation (2) and the mass of the beam  $M = \rho AL$ . The resulting characteristic equation for the wave-number  $k$

$$1 + \cos(k_n L) \cosh(k_n L) = \frac{m}{M} k_n L (\cos(k_n L) \sinh(k_n L) - \sin(k_n L) \cosh(k_n L)) \quad (7)$$

obviously reduces to the usual fixed-free cantilever characteristic equation in the limit  $m/M \rightarrow 0$ . From the boundary condition  $-EI\mathbf{u}'' = \mathbf{0}$  at  $x = L$ ,

$$A_2 = -\frac{\cos(k_n L) + \cosh(k_n L)}{\sin(k_n L) + \sinh(k_n L)} A_1 \quad (8)$$

and consequently

$$\chi_n = A_n \left( \cos(k_n x) - \cosh(k_n x) - \frac{\cos(k_n L) + \cosh(k_n L)}{\sin(k_n L) + \sinh(k_n L)} \times (\sin(k_n x) - \sinh(k_n x)) \right) \quad (9)$$

<sup>d</sup>For a generic tip force of the form  $\mathbf{f} = C(k)\mathbf{u}|_{x=L}$  the characteristic equation is:

$$1 + \cos(k_n L) \cosh(k_n L) = \frac{C}{EIk^3} (\cos(k_n L) \sinh(k_n L) - \sin(k_n L) \cosh(k_n L))$$

A mass (or molecule) attached by an elastic bond to the tip of the CNT is of this form but involves an additional degree-of-freedom and attendant governing equation for its dynamics.

which is the same as fixed-free solution<sup>31,32</sup> since it does not depend on the tip mass  $m$ . For  $m/M = 0$ , the roots of (7) are  $k_n L \approx \pi\{0.596864, 1.49418, 2.50025, 3.49999, n + 1/2\}$ ; whereas, for  $m/M = 1$ , the roots are  $k_n L \approx \pi\{0.397224, 1.283151, 2.270865, 3.264784, \dots\}$ , and for  $m/M \rightarrow \infty$ , the roots are  $k_n L \approx \pi\{1.24988, 2.25, 3.25, \dots\}$ , see Ref. [33] and Figure 3. In fact, the first frequency of the fixed-free beam with a tip mass converges to zero and the second becomes the fundamental (first mode) of a fixed-pinned beam ( $\mathbf{u} = \mathbf{u}'' = \mathbf{0}$  at  $x = L$ ) in the limit  $(m/M) \rightarrow \infty$ .

An orthogonality condition

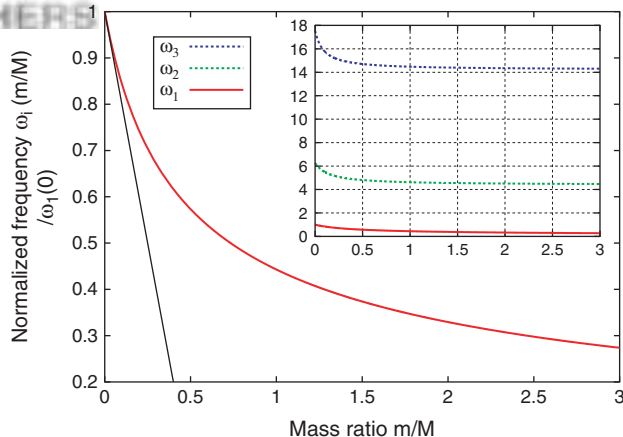
$$(\omega_i^2 - \omega_j^2) \left[ \frac{1}{L} \int_0^L \chi_i \chi_j dx - \frac{EI}{\rho AL} \left[ \chi_i \chi_j''' - \chi_j \chi_i''' + \chi_i' \chi_j'' - \chi_j' \chi_i'' \right] \Big|_{x=L} \right] = (\omega_i^2 - \omega_j^2) \left[ \frac{1}{L} \int_0^L \chi_i \chi_j dx + \frac{m}{M} \chi_i \chi_j \Big|_{x=L} \right] = 0 \quad (10)$$

is obtained from the difference in modes substituted into governing Eq. (1). To facilitate the subsequent analysis, we impose the normalization

$$\frac{1}{L} \int_0^L \chi_i \chi_j dx + \frac{m}{M} \chi_i \chi_j \Big|_{x=L} = \delta_{ij} \quad (11)$$

where  $\delta_{ij}$  is the Kronecker delta. It is important to note that this normalization is only valid for frequencies that satisfy the characteristic Eq. (7). Consequently, the potential energy due to bending can be reduced to:

$$U(t) = \frac{1}{2} \int_0^L EI \mathbf{u}'' \cdot \mathbf{u}'' dx = \frac{1}{2} \left( \int_0^L EI \mathbf{u}'''' \cdot \mathbf{u} dx + [EI \mathbf{u}''' \cdot \mathbf{u} - EI \mathbf{u}'' \cdot \mathbf{u}']_0^L \right)$$



**Fig. 3.** The inset shows the change in the three lowest frequencies of an Euler–Bernoulli beam with the rigid attachment of a tip mass. The fundamental  $\omega_0$ , shown in both the inset and the larger graph, converges to zero and the second harmonic converges to the fundamental of a fixed-pinned beam. Note the black line in the main plot is the linearization  $\omega_0(m/M)/\omega_0(0) = 1 - 2(m/M)$ .

$$\begin{aligned}
 &= \frac{EIL}{2} \sum_{n,m} k_n^4 \left( \frac{1}{L} \int_0^L \mathbf{u}_n \cdot \mathbf{u}_m \chi_n \chi_m dx \right. \\
 &\quad \left. + \frac{m}{M} [\mathbf{u}_n \cdot \mathbf{u}_m \chi_n \chi_m]_{x=L} \right) \\
 &= \frac{EIL}{2} \sum_{n,m} k_n^4 \mathbf{u}_n(t) \cdot \mathbf{u}_m(t) \\
 &\quad \times \left( \frac{1}{L} \int_0^L \chi_n \chi_m dx + \frac{m}{M} [\chi_n \chi_m]_{x=L} \right) \\
 &= \frac{EIL}{2} \sum_n k_n^4 \|\mathbf{u}_n(t)\|^2 \left( \frac{1}{L} \int_0^L \chi_n^2 dx + \frac{m}{M} [\chi_n^2]_{x=L} \right) \\
 &= \frac{EI}{2} \sum_n k_n^4 \|\mathbf{u}_n(t)\|^2 \quad (12)
 \end{aligned}$$

using (11). At equilibrium, equipartition of the potential and kinetic energies [Ref. [34], Section 6.4] requires

$$\langle E_n \rangle = 2 \langle U_n \rangle = k_B \theta \quad (13)$$

where  $E_n$  is the total energy,  $U_n = (EI/2)k_n^4 u_n^2$  is the potential energy per mode and per transverse coordinate,  $k_B$  is the Boltzmann constant and  $\theta$  is the temperature. (The angle brackets here and later denote a long-time average.) Consequently,

$$\langle u_n^2 \rangle = \frac{k_B \theta}{EIL k_n^4} = \frac{k_B \theta}{M \omega_n^2} \quad (14)$$

where  $u_n$  represents either of the components of  $\mathbf{u}_n$  since they are independent and equivalent. Thus the often cited result derived by Treacy et al.<sup>8,35</sup> generalized to a CNT with an added mass at the tip is

$$\begin{aligned}
 |u|_L^2 &\equiv \langle u^2|_L \rangle = \left\langle \sum_{n=0}^{\infty} u_n \chi_n|_L \right\rangle^2 = \sum_{n=0}^{\infty} \langle u_n^2 \rangle |\chi_n|_L^2 \\
 &= \frac{k_B T}{M} \sum_{n=0}^{\infty} \omega_n^{-2} |\chi_n|_L^2, \quad (15)
 \end{aligned}$$

assuming that the modes are incoherent (i.e., have random phases) so that  $\langle u_n u_m \rangle = 0$  if  $n \neq m$ . Here,  $|\chi_n|_L^2 \equiv \chi_n^2(x=L)$  is given by

$$\frac{m}{M} |\chi_n|_L^2 = 1 - \frac{1}{L} \int_0^L \chi_n^2 dx \quad (16)$$

which is a consequence of the normalization (11).<sup>e</sup>

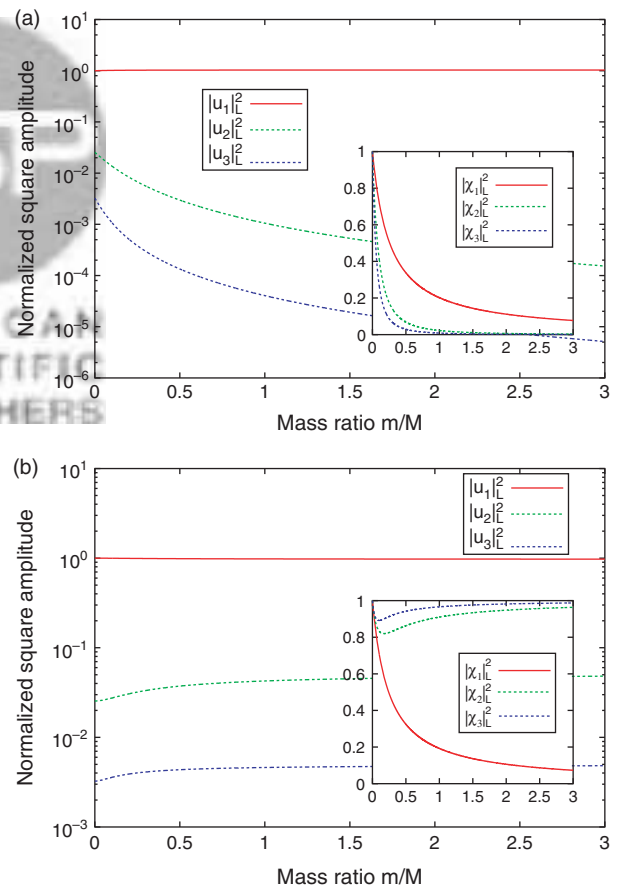
There is obviously a lower limit to the applicability of the continuum approximation for a discrete system like a CNT but the series (15) converges quickly. In fact,

$$\sum_{n=0}^{\infty} (k_n L)^{-4} = \frac{\rho A}{EIL^4} \sum_{n=0}^{\infty} \omega_n^{-2} = \frac{1}{12} \quad (17)$$

<sup>e</sup>The ratio  $|\chi_n|_L^2$  to  $(1/L) \int_0^L \chi_n^2 dx$  is equal to 4 at  $m=0$  as mentioned in the Introduction. Krishnan et al.<sup>35</sup> takes  $A_n = 1/2$  so that  $\chi_L = 1$  for a CNT without an attached molecule.

for a cantilever with  $m/M = 0$  given the roots of the characteristic Eq. (7). The ratio of the contribution of the first mode to the second is  $u_1/u_0 = (k_0/k_1)^4$  which is approximately 0.025 for an unencumbered cantilever. The ratio of the contribution of the first mode to all the rest is  $(1/u_0) \sum_{n=1}^{\infty} u_n = (1/k_0^4) \sum_{n=1}^{\infty} k_n^{-4}$  which is approximately 0.030 for an encumbered cantilever. So it is apparent that the tip trajectory of a thermally excited beam (with a small tip mass or without) can be accurately approximated by a single harmonic oscillator with natural frequency  $\omega_0$ .

Since the relation (15) depends on the solution of the characteristic Eq. (7), there is no closed form solution to the influence of added mass on the tip displacement amplitude. Figure 4 displays the numerical solution to amplitude response of the beam as a function of the attached mass. Figure 4a shows that the fundamental mode, which contributes most of the tip amplitude, is relatively insensitive to added mass whereas the higher modes decrease with added mass. The fact that  $|u_0|^2$  is essentially constant implies that  $u_L^2 \sim k^4$ . This same trend for the fundamental



**Fig. 4.** The change in the equilibrium tip displacement  $|u|_L^2$  and average displacement  $\|\mathbf{u}\|^2$  of an Euler–Bernoulli beam with the rigid attachment of a tip mass. The inset shows the modal contributions to the tip displacement  $|\chi_i|_L^2$  and  $\|\chi\|_L^2$  as a function of  $m/M$ .



mode can be seen in Figure 4(b) which shows the average squared displacement along the length of the beam to be

$$\|u_n\|^2 = \frac{k_B \theta}{EIL} k_n^{-4} \|\chi_n\|^2 \quad (18)$$

where  $\|\chi_n\|^2 \equiv (1/L) \int_0^L \chi_n^2 dx$ . The figure also shows that the contribution of the higher modes increases with  $m$ . Recalling the fact that  $\omega_0 \rightarrow 0$  in the limit  $(m/M) \rightarrow \infty$ , i.e., the first mode becomes static, it is not clear what happens to equipartition condition which stipulates  $k_B T = U_0 = M \omega_0^2 \langle u_0^2 \rangle$ . The fact that the response of a beam with a tip mass converges to that of a fixed-pinned beam is also demonstrated by the fact that the higher modes contributions to the tip displacement decreases while the average displacement increases, i.e., the maximum displacement amplitude moves away from the tip. Practically speaking, for a beam with a small mass attached the limit  $(m/M) \rightarrow \infty$  is not particularly relevant other than as a check on the theory, while the insensitivity of the tip displacement to added mass makes equilibrium amplitude detection of attached mass impractical.

## 2.2. Langevin Model

Although the beam theory is very useful in describing some aspects of the behavior of the CNT, the observation that approximately 97% of the motion in thermal equilibrium is due to the fundamental mode allows treatment of the tip displacement as a simple noisy harmonic. More precisely, we employ a Langevin oscillator to represent the fundamental mode in a bath of a finite number of higher modes and the attendant spreading of the fundamental peak from an undamped resonance displayed by the MD data in Figure 2. For simplicity, throughout this section we use the notation  $u$  for the tip displacement  $u_L$ .

The Langevin equation is a stochastic ordinary differential equation (SODE) that can take the form of a damped harmonic oscillator

$$m\ddot{u} = -ku - \gamma\dot{u} + r \quad (19)$$

driven by a random force  $r = r(t)$  representing the higher modes. Here,  $m$  is an (effective) mass,  $k$  is an (effective) spring constant and  $\gamma$  is a damping coefficient.<sup>f</sup> A more formally correct version of (19) is

$$m dv = -(ku + \gamma v) dt + \sigma dB_t \quad (20)$$

where  $v = du$ , see Ref. [36, Chapters 3 & 5]. The Wiener process representing the white noise  $r$  is partially defined by  $dB_t \sim \mathcal{N}(0, t) = \sqrt{t} \mathcal{N}(0, 1)$ , so that

$$r dt = \sigma dB_t = \sigma \sqrt{dt} w_t \quad (21)$$

<sup>f</sup>A Langevin system with multiple resonances could also be developed but this level of complexity is unnecessary since we confine our attention to frequencies near the fundamental resonance. Without higher resonances, decay of the amplitude spectrum far above resonance is like  $\omega^{-4}$  which deviates significantly from the observed  $\omega^{-2}$  behavior.<sup>30</sup>

for  $w_t \sim \mathcal{N}(0, 1)$  and  $\mathcal{N}(0, 1)$  being the zero mean, unit variance normal distribution. Following Kubo [37, Chapter 1], a Fourier transform of this SODE leads to the spectral intensity of the response  $\mathcal{F}_u$  being related to the intensity of the random force  $\mathcal{F}_r$  by

$$\mathcal{F}_u(\omega) = |H(\omega)|^2 \mathcal{F}_r(\omega) \quad (22)$$

where the magnitude of the system response  $H(\omega)$  is

$$|H(\omega)|^2 = (m^2[(\omega_0^2 - \omega^2)^2 + (\omega_0 \omega)^2 Q^{-2}])^{-1} \quad (23)$$

and the natural frequency is  $\omega_0^2 = k/m$ . From (23), the frequency at maximum amplitude  $\varpi$  is

$$\frac{\varpi}{\omega_0} = \sqrt{1 - \frac{1}{2Q^2}} \quad (24)$$

where the quality factor  $Q$  is defined  $Q = (m\omega_0)/\gamma$ . So for high  $Q$ ,  $\varpi \approx \omega_0$  and  $|H(\varpi)|^2 \approx |H(\omega_0)|^2 = Q^2/(m^2\omega_0^2)$ . The quality factor also determines the (velocity) relaxation time  $m/\gamma = Q/\omega_0$ .

As in Ref. [37, Chapter 1], we make the assumption that the random force is uncorrelated so that  $\mathcal{F}_r(\omega) = I_r$ . The auto-correlation

$$\mathcal{A}_r(t) = \langle r(0)r(t) \rangle = 2\pi I_r \delta(t) \quad (25)$$

results from Wiener-Khinchin theorem where  $\delta(t)$  is the Dirac delta.<sup>g</sup> Consequently,

$$\begin{aligned} \mathcal{A}_u(t) &= \mathcal{F}_{\omega \rightarrow t}^{-1} |H(\omega)|^2 I_r \\ &= \frac{\pi I_r}{m\gamma\omega_0^2} \left( \cos(\omega_d t) + \frac{\gamma}{2\omega_d} \sin(\omega_d t) \right) \exp\left(-\frac{\gamma t}{2}\right) \end{aligned} \quad (26)$$

where  $\omega_d^2 = \omega_0^2 (1 - (1/2Q)^2)$  and  $\mathcal{F}_{\omega \rightarrow t}^{-1}$  is the inverse Fourier transform. Note that  $\mathcal{A}_u$  is continuous, unlike  $\mathcal{A}_r$ . Now using equipartition  $\langle U_0 \rangle = \langle \frac{1}{2} m \omega_0^2 u^2 \rangle = \frac{1}{2} k_B \theta$ , i.e., assuming the system is in equilibrium,

$$\langle u^2 \rangle = \mathcal{A}_u(0) = \frac{k_B \theta}{m\omega_0^2} = \frac{\pi I_r}{m\gamma\omega_0^2}. \quad (27)$$

Consequently,

$$\pi I_r = \gamma k_B \theta = \frac{m\omega_0}{Q} k_B \theta \quad (28)$$

where we have approximated the kinetic energy as  $\frac{1}{2} m \omega_0^2$ , which is valid for high  $Q$ . This is a similar result that developed for beam theory (14) in the previous section. The (equilibrium) fluctuation-dissipation theorem also results

$$\langle r(0)r(t) \rangle = \gamma k_B \theta \delta(t) \quad (29)$$

<sup>g</sup>The theorem states that intensity  $\mathcal{F}_u$  is related to the auto-correlation  $\mathcal{A}_u = \langle u(0)u(t) \rangle$  through the Fourier transform  $\mathcal{F}_u(\omega) = \mathcal{F}_{t \rightarrow \omega} \mathcal{A}_u(t)$  for a stationary process.

giving  $\sigma^2 = \pi I_r = \gamma k_B \theta$ . We now can obtain the peak amplitude from the system response from (22)

$$I_u(\omega) = |H(\omega)|^2 I_r = \frac{\omega_0 k_B \theta}{\pi m Q [(\omega_0^2 - \omega^2)^2 + \omega^2 \omega_0^2 Q^{-2}]} \quad (30)$$

so that

$$I_u(\omega_0) = \frac{Q k_B \theta}{\pi m \omega_0^3} \quad (31)$$

To discern the intrinsic fundamental parameters, we non-dimensionalize (19)

$$\ddot{u} = -\tilde{u} - \frac{1}{Q} \dot{\tilde{u}} + \tilde{r} \quad (32)$$

where  $\tilde{u} = u/\ell$ ,  $\tilde{r} = r/(m\omega_0^2\ell)$ ,  $\tilde{t} = t\omega_0$ , and  $\ell = \sqrt{k_B\theta/m\omega_0^2}$ . Now (21) becomes

$$\tilde{r} d\tilde{t} = \frac{r}{m\omega_0\ell} dt = \frac{\sigma\sqrt{dt}}{m\omega_0\ell} w_t = \frac{1}{Q} \sqrt{Q d\tilde{t}} w_t \quad (33)$$

and (20) becomes

$$dv = -u dt - \frac{1}{Q} (v dt - \sqrt{Q dt} w_t) \quad (34)$$

where we have dropped the tildes for clarity. This model of the white noise process is amenable to numerical solution and retains the property  $dB_t = \sqrt{dt}$ . It takes the variance of the non-dimensional white noise to be  $\sigma^2 = dt/Q$  and the non-dimensional maximum intensity as  $\mathcal{F}_u(\omega_0) = Q/\pi$ .

### 3. FILTERING ALGORITHMS

The previous section developed the characteristic Eq. (7) that gave the sensitivity of changes in frequency to added mass. It should be noted that this was not strictly an equilibrium result. On the other-hand, the Treacy equipartition argument is necessary for the amplitude change estimate, and shows that the amplitude is largely insensitive to small attached masses. Clearly, the sensitivity of the frequency of the primary resonance recommends it over a scheme based on detecting changes in amplitude. Nevertheless, the changes in frequency are potentially very small for  $m \ll M$  relative to the thermal noise (which was directly related to the quality factor through the fluctuation-dissipation theorem in the previous section).

Our goals require identifying the change in the fundamental frequency of oscillation of a CNT-based sensor over times shorter than the residence times of the attached molecules.<sup>h</sup> Unfortunately, identifying frequency shifts using traditional spectral methods, such as the widely-employed fast Fourier transform (FFT), have several drawbacks pertinent to achieving this objective. The primary

issue being that sampling over a long period of time is required to obtain the necessary resolution in frequency space. This fact alone recommends the application of a real-time estimate of the maximum amplitude frequency via techniques like the Kalman filter.

In this section we first review the basics of spectral analysis and then develop a specialized Kalman filter (KF) from the standard formulation. The Kalman filter is well-known as the optimal filter for linear dynamical systems with Gaussian process and measurement noise. An array of these specialized KFs with an arbitrary frequency spacing can be employed in parallel, together with an adaptive algorithm developed in this section, to track the peak intensity of a resonant signal, as we demonstrate in Section 4.

#### 3.1. Fourier Analysis

The discrete time Fourier transform (DTFT) extracts spectral information from a discrete time signal  $u_{t_j}$

$$\mathbb{F}_{i \rightarrow \omega} u_{t_j} = \sum_{n=-\infty}^{\infty} u_n \exp(in\omega) \quad (35)$$

where, in this context,  $u_{t_j}$  is  $u(t)$  is sampled on period  $\Delta t$  (and not the  $j$ -th). It is in contrast to the discrete Fourier transform (DFT) (and its implementation the FFT)

$$\mathbb{F}_{i \rightarrow k} u_n = \sum_{n=0}^{N-1} u_n \exp\left(i \frac{2\pi}{N} nk\right) \quad (36)$$

which assumes a periodic extension of the signal and only provides spectral content at discrete frequencies spaced  $\Delta\omega = 2\pi/T$ , where  $T = N\Delta t$ . Both are connected to the Fourier transform by the approximate relationships:  $\mathcal{F}_{i \rightarrow \omega} u(t) \approx \mathbb{F}_{j \rightarrow \omega} u_{t_j} \Delta t$ , and  $\mathcal{F}_{i \rightarrow k} u(t) \approx \mathbb{F}_{j \rightarrow k} u_{t_j} \Delta t$  for a periodic signal.

Since only we only observe for a finite time  $T$ , the DTFT of a sinusoidal signal  $u(t) = \sin(\omega_0 t)$  is equivalent to the Fourier transform of a windowed sine

$$\begin{aligned} \mathcal{F}_{i \rightarrow \omega}(\sin(\omega_0 t) h(t)) &= \mathcal{F}_{i \rightarrow \omega} \sin(\omega_0 t) * \mathcal{F}_{i \rightarrow \omega} h(t) \\ &= \frac{iT}{2\sqrt{2}\pi} \left( \text{sinc}\left(\frac{T}{2}(\omega + \omega_0)\right) - \text{sinc}\left(\frac{T}{2}(\omega - \omega_0)\right) \right) \end{aligned} \quad (37)$$

where  $\text{sinc}(x) = \sin(x)/x$  is the sinc function, the “top hat” window  $h(t) = \vartheta(t - T/2) - \vartheta(t + T/2)$  is composed of two Heaviside functions  $\vartheta(t)$  and  $*$  is the convolution operator.<sup>i</sup> The sinc function  $\text{sinc}((T/2)(\omega + \omega_0))$  has lobes

<sup>i</sup>Strictly speaking

$$\begin{aligned} \mathcal{F}_{i \rightarrow \omega}(\sin(\omega_0 t) h(t)) &= \frac{iT}{2\sqrt{2}\pi} \left( \text{sinc}\left(\frac{T}{2}(\omega + \omega_0)\right) - \text{sinc}\left(\frac{T}{2}(\omega - \omega_0)\right) \right) \\ &\quad + \sqrt{\frac{\pi}{2}} \delta(\omega + \omega_0) \sin\left(\frac{T}{2}(\omega + \omega_0)\right) - \sqrt{\frac{\pi}{2}} \delta(\omega - \omega_0) \sin\left(\frac{T}{2}(\omega - \omega_0)\right), \end{aligned}$$

where the last two terms are not observed in finite precision experiments nor simulations.

<sup>h</sup>Ultimately, the detection time is limited by the physical relaxation time of the CNT to mechanical perturbations, which also can be directly related to the quality factor of the CNT.

with zeros at  $2\pi n/T + \omega_0$ ,  $n = \dots, -2, -1, 1, 2, \dots$ , which decrease in width as  $T \rightarrow \infty$ .

Parseval's theorem  $\sum_{n=0}^{N-1} u_n^2 = (1/N) \sum_{k=0}^{N-1} |\mathbb{F}_{n \rightarrow k} u_n|^2$  provides a means of the normalization of the spectrum through the average power, e.g., for the sine

$$\begin{aligned}\bar{\mathcal{P}}_u &= \frac{1}{T} \int_0^T u^2(t) dt = \frac{1}{T} \left( \frac{T}{2} - \frac{\sin(T\omega_0)}{\omega_0} \right) \\ &= \frac{1}{2} - \text{sinc}(T\omega_0) \approx \frac{1}{N} \sum_{j=0}^{N-1} u_{t_j}^2\end{aligned}\quad (38)$$

to arrive at an intensity  $\mathcal{J}_u$ . We recognize the first term as  $\pi/(T\Delta\omega)$  and the second as a perturbation due to not always observing an integral number of periods of the sine. Defining estimates of the spectral intensity of the signal  $u_{t_j}$  as

$$\begin{aligned}\mathcal{J}_u(\omega) &\approx \lim_{T \rightarrow \infty} \frac{1}{2\pi T} |\mathbb{F}_{j \rightarrow \omega} u_{t_j}|^2 \Delta t^2 \\ &= \lim_{N \rightarrow \infty} \frac{1}{2\pi N} |\mathbb{F}_{j \rightarrow \omega} u_{t_j}|^2 \Delta t\end{aligned}\quad (39)$$

and likewise

$$\mathcal{J}_u(k\Delta\omega) \approx \lim_{N \rightarrow \infty} \frac{1}{2\pi N} |\mathbb{F}_{j \rightarrow k} u_{t_j}|^2 \Delta t, \quad (40)$$

we see that these definitions are consistent with Parseval's theorem

$$\begin{aligned}\int_{-\infty}^{\infty} \mathcal{J}_u(\omega) d\omega &\approx \sum_{j=0}^{N-1} \mathcal{J}_u(j\Delta\omega) \Delta\omega \\ &= \lim_{N \rightarrow \infty} \frac{1}{2\pi N} \sum_{j=0}^{N-1} |\mathbb{F}_{j \rightarrow \omega} u_{t_j}|^2 \Delta t \frac{2\pi}{N\Delta t} \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=0}^{N-1} u_{t_j}^2 = \bar{\mathcal{P}}_u\end{aligned}\quad (41)$$

In the following we employ the notation

$$u^2(\omega) = \frac{1}{N} |\mathbb{F}_{j \rightarrow \omega} u_{t_j}|^2 \quad (42)$$

and refer to it as the “power” following the signal processing literature.

Unfortunately, direct application of the DTFT to a finite set of frequencies is difficult in this application given the wide array of challenges that must be met. The greatest challenge is minimizing the computational cost, which also implies being parsimonious with the frequencies that are included in the analysis and the communication between the filters that track them. This last constraint effectively eliminates the viability of the standard FFT. Limiting the interaction between different frequency estimates is necessary for scanning each of  $N$  frequencies in parallel. Also, in order to perform real time detection of finite-duration

events, methods that utilize time averages are inappropriate because of their long response time to the transient. In conjunction with this observation, the limited number of frequencies that can be included means that the frequencies being analyzed must change over time as the method converges. Facilitating such an approach requires the ability to transfer information from a previous set of frequencies to a new set in order to perform rapid restarting to quickly identify the next best estimate. Given the frequency content of the CNTs being in the GHz and higher, it is also necessary that the method be amenable to sampling at a sample rate below the frequencies being considered (i.e., not subject to a Nyquist restriction), and to be robust in the presence of noise caused by the higher order modes of the CNT.

There is a large body of literature surrounding alternative methods to estimate the FT at a given frequency in real time, and we will not attempt to exhaustively review the literature in this area. Rather, we point to two main approaches: recursive Fourier transforms (RFTs) and sequential Bayesian analysis applied to Fourier coefficient estimation.

Bayesian sequential analysis is the application of Bayesian theory to sequentially-arriving data with the objective being estimation of a function of this data. Estimation of Fourier coefficients performed with this method, e.g., Ref. [38], meets many of the required constraints. However, estimating the posterior distribution is expensive when the set of frequencies is close together and does not span a significant extent of frequency space. A notion embedded within the Bayesian approach involves selecting an optimal frequency among a set of frequencies based on an estimate of each's likelihood of being the correct frequency. This general notion will prove useful in this work, although more efficient methods will be used that do not require the communication necessary in the Bayes framework.

RFTs work in either the mode of a fixed number of data points with recursive addition of new data and removal of old data, e.g., Ref. [39], or by using a recursive least squares solution to update the estimate, e.g., Ref. [40]. Recent improvements have removed the restriction of a uniform sampling rate and arrived at a convenient update formula [Ref. [41], Eq. (16)]. This formula bears a number of similarities to the one derived in the next section. A principle difference, both in the approach and the final update equation, is the treatment of noise. In the overarching least-squares framework of Ref. [41] measurement noise can enter through a weighting matrix, which, for example, can be used to reduce the contribution of old data to the update to produce a windowing effect. However, in order to reduce the cost of the method the weighting matrix was required to be the identity in the cited work, meaning that uncertainty could not be directly incorporated into the frequency estimate. Our formulation will explicitly



account for the error in the system using an approach based on Kalman filtering, as derived in the next section. This filtering will form the basis of the inner iteration, with a closed-loop outer iteration designed to converge to the peak-amplitude frequency. As previously mentioned, the amount of literature in this area is vast, making a full survey infeasible. However, the authors are unaware of this inner/outer loop approach to frequency estimation being used in other applications.

### 3.2. Kalman Filter

As an alternative to traditional Fourier transform approaches, this work considers an approach based on Kalman filtering. The Kalman filter has been in use for almost 50 years<sup>42</sup> in a wide variety of applications. It provides an optimal estimate of the state of a linear, time-varying system with Gaussian process and measurement noise. For the present application, its key advantages are:

- (a) its natural incorporation of noise which can be augmented with the theoretical estimates developed in the previous sections,
- (b) the fact that it is not tied to a particular sample rate, and
- (c) its formulation in terms of the present state and sample only, leading to an incremental cost to an improved estimate over time.

An approximation of the KF particular to our application will be derived later in this section that can easily be restarted and interpolated assuming a continuous Fourier transform in order to facilitate inner loop/outer loop frequency estimation. Most importantly, the computational cost of each KF is very low and each frequency can be considered individually within the inner, Kalman-based loop.

Kalman filtering in typical applications involves estimating the state of the system in the presence of additive process and measurement noise. The estimate is optimal for linear systems with Gaussian white noise. Estimation of spectral content of a signal is a slightly different application. For the present theoretical treatment, we consider there to be no significant noise associated with taking a measurement although future can incorporate it into the Kalman filter framework. Rather, the uncertainty will be associated with the motions of higher order modes relative the fundamental mode. As a model, consider a system composed of  $M$  harmonic components so there are  $2M$  degrees of freedom (position and velocity for each oscillator) undergoing the following dynamics without process noise:

$$\dot{x}_k^M = F_k^M x_{k-1}^M, \quad F_k^M = \begin{bmatrix} 0 & 1 \\ -K & 0 \end{bmatrix}, \quad (43)$$

where  $I$  is an  $M \times M$  identity matrix and  $K$  is an  $M \times M$  diagonal, positive-definite matrix. The observation procedure also is noise-free

$$y_k^M = H_k^M x_k^M, \quad H_k^M = [1 \dots 1 \ 0 \dots 0], \quad (44)$$

with the change from 1 to 0 occurring from the  $M$  to  $M + 1$  entry to sum over all the positions. Hence, the observation is

$$y_k^M = x_k^1 + \sum_{m=2}^M x_k^m, \quad (45)$$

where  $x_k^1$  contains all the amplitude information associated with the fundamental frequency of oscillation. The theoretical development of the previous sections suggest that  $\lim_{M \rightarrow \infty} \sum_{m=2}^M x_k^m = W_k$ , where  $W_k$  is a white noise process of appropriate amplitude if  $\langle (x_k^1)^2 \rangle \gg \lim_{M \rightarrow \infty} \sum_{m=2}^M \langle (x_k^m)^2 \rangle$ . The final observation process used in this work can then be written as

$$y_k \approx x_k^1 + W_k. \quad (46)$$

by having the noise term account of the sum of the motions of the all the modes above the fundamental.

In this model, the oscillators are uncoupled, allowing the position of each to be spectrally decomposed as:

$$x_k^i(t) = a_i \cos \omega_i t + b_i \sin \omega_i t. \quad (47)$$

Taking the unknowns of interest to be  $a_1$  and  $b_1$  and letting  $x$  denote this new state vector, the Kalman filtering problem can be written as:

$$\begin{aligned} x_k &= F_k x_{k-1} \\ y_k &= H_k x_k + W_k. \end{aligned} \quad (48)$$

In this model, the noise  $W_k$  is Gaussian with covariance matrix  $R_k$ . The Kalman filter (KF) provides an estimate  $\hat{x}_k$  with minimum covariance  $P_k$  and is the optimal linear estimate given this noise. To filter the CNT tip position, the state would typically be considered to be one of transverse displacements. However, our intended purpose is to estimate frequency content so the state  $x_k$  is instead taken to be the Fourier coefficients of this signal. Within the context of estimating the spectral content in a signal  $u_i$ , we use the following values for the states and matrices:

$$\begin{aligned} x_k &= \begin{bmatrix} a_k \\ b_k \end{bmatrix}, \quad F_k = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ H_k^T &= \begin{bmatrix} \cos \omega t \\ \sin \omega t \end{bmatrix}, \quad R_k = \sigma_k^2. \end{aligned} \quad (49)$$

so that  $\|x\|^2$  estimates  $u^2(\omega) \equiv (1/N) |F_{i \rightarrow \omega} u_{t_i}|^2$ . While in this case development has focused on noise sources arising from higher order modes, it is also applicable to more general noisy measurements of system parameters. For example, the Langevin model presented in this work does not

precisely conform to this structure because the white noise source embedded in this model is shaped by the system response and is integrated in time to obtain the observational uncertainty. As will be shown in the Results section, the proposed method is applicable to this situation as well.

For the CNT, theoretical estimates have been made in this work and in Ref. [8] that enable determination of the various uncertainties present. In particular, the observation noise is attributed to the interaction of the high frequency modes with the fundamental (as in the Langevin model of Section 2.2). The variance of this noise can be estimated based on the amount of energy in these modes. Similarly, the energy in the fundamental model can be related to the amplitude of oscillation of the CNT. From this amplitude, the initial variance can be determined assuming the phase is random. We employ the estimates:

$$c_0 = \langle u_1^2 \rangle \approx 0.97 \langle u^2 \rangle \quad (50)$$

$$\sigma_0^2 = \sum_{i=2}^n \langle u_i^2 \rangle \approx 0.03 \langle u^2 \rangle \quad (51)$$

re-derived from beam theory in Section 2 for a CNT without a tip mass. Alternately, we could have initialized the covariance with the estimate  $\sigma^2 \sim Q^{-1}$  from the Langevin model but chose not to given that  $Q$  is not typically well-characterized.

The Kalman update is composed of three steps: a predictor, a measurement, and a corrector. The predictor step is

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1} = x_{k-1} \quad (51)$$

$$P_{k|k-1} = F_k P_{k-1} F_k^T = P_{k-1}. \quad (52)$$

In this case, the predictor is trivial due to the assumption of stationary amplitudes and phases, i.e., we assume a so-called wide-sense stationary process in (49). The measurement is

$$y_k = z_k - H_k \hat{x}_{k-1}, \quad (53)$$

which is again straightforward. Note the measurement  $y_k$  is the difference between the actual system observation  $z_k$  and the appropriately weighted sum of the cosine and sine.

Finally, we consider the standard Kalman corrector step

$$S_k = H_k P_{k-1} H_k^T + R_k \quad (54)$$

$$K_k = P_{k-1} H_k^T S_k^{-1} \quad (55)$$

$$\hat{x}_k = \hat{x}_{k-1} + K_k y_k \quad (56)$$

$$P_k = (I - K_k H_k) P_{k-1}. \quad (57)$$

Note in this formulation, the innovation covariance  $S_k$  is a scalar. This step is non-trivial, with the most significant complication being the arbitrariness of  $P_{k-1}$  because the initial condition,  $P_0$ , must be assumed. Understanding and improving the method relies on understanding  $P_{k-1}$  and

the matrix-vector product  $P_{k-1} H_k^T$ . To accomplish this, first consider a form of the matrix  $P$ , denoted  $P^\dagger$ , to be

$$P_{k-1}^\dagger = c_{k-1} \begin{bmatrix} \cos(\omega t_k) H_k^T & \sin(\omega t_k) H_k^T \end{bmatrix} \\ \equiv c_{k-1} \begin{bmatrix} \cos(\omega t_k) H_k \\ \sin(\omega t_k) H_k \end{bmatrix} = c_{k-1} H_k^T H_k. \quad (58)$$

This makes the product

$$P_{k-1}^\dagger H_k^T = c_{k-1} \begin{bmatrix} \cos(\omega t_k) |H_k|^2 \\ \sin(\omega t_k) |H_k|^2 \end{bmatrix} \\ = c_{k-1} \begin{bmatrix} \cos(\omega t_k) \\ \sin(\omega t_k) \end{bmatrix} \equiv c_{k-1} H_k^T. \quad (59)$$

Upon substitution, the following simplifications arise:

$$S_k^\dagger = c_{k-1} + \sigma_k^2 \quad (60)$$

$$K_k^\dagger = \frac{c_{k-1}}{c_{k-1} + \sigma_k^2} H_k^T. \quad (61)$$

The update for  $P_k^\dagger$  is then

$$P_k^\dagger = \left( 1 - \frac{c_{k-1}}{c_{k-1} + \sigma_k^2} \right) P_{k-1}^\dagger, \quad (62)$$

where we have used the fact that

$$H_k^T H_k P_{k-1}^\dagger = P_{k-1}^\dagger. \quad (63)$$

Equation (62) provides the decay rate of the covariance matrix since

$$c_k = \left( 1 - \frac{c_{k-1}}{c_{k-1} + \sigma_k^2} \right) c_{k-1}. \quad (64)$$

Using this covariance matrix also enables deeper understanding of the evolution of the estimate. The update equation is

$$\hat{x}_k^\dagger = \hat{x}_{k-1}^\dagger + \frac{c_{k-1}}{c_{k-1} + \sigma_k^2} H_k^T y_k. \quad (65)$$

This shows that the noise level adjusts the gain into which the errors in the observed and predicted signals are added into the harmonics through the term  $c_{k-1}/c_{k-1} + \sigma_k^2$ . This leads to the expectation that the Kalman response at any frequency is equal to a top hat function at the origin convolved with the true signal after the covariance converges, as in Eq. (37).

While illustrative, there is a clear contradiction between Eqs. (62) and (58). The latter shows that  $P_k^\dagger$  is of a constant form with magnitude decay while the former shows that the matrix itself changes as the harmonics evolve. This contradiction can be resolved by choosing a perturbation to  $P_k^\dagger$ ,

$$P_{k-1} = c_{k-1} \begin{bmatrix} \cos(\omega t_{k-1}) H_{k-1}^T & \sin(\omega t_{k-1}) H_{k-1}^T \end{bmatrix}. \quad (66)$$

For small values of  $\omega\Delta t$ , we can use Taylor series expansion about  $t_{k-1} = t_k - \Delta t$  to obtain

$$\begin{aligned} P_{k-1} H_k^T &= c_{k-1} \begin{bmatrix} \cos^2(\omega t_{k-1}) \cos(\omega t_k) \\ + \cos(\omega t_{k-1}) \sin(\omega t_{k-1}) \sin(\omega t_k) \\ \cos(\omega t_{k-1}) \sin(\omega t_{k-1}) \cos(\omega t_k) \\ + \sin^2(\omega t_{k-1}) \sin(\omega t_k) \end{bmatrix} \\ &+ O(\Delta t^2) \\ &\approx c_{k-1} H_k^T + \omega \Delta t \begin{bmatrix} 2 \cos^2(\omega t_k) \sin(\omega t_k) + \sin^3(\omega t_k) \\ - \cos^2(\omega t_k) \sin(\omega t_k) \\ \cos(\omega t_k) \sin^2(\omega t_k) - \cos^3(\omega t_k) \\ - 2 \cos(\omega t_k) \sin^2(\omega t_k) \end{bmatrix} \\ &= c_{k-1} H_k^T + c_{k-1} \Delta t \dot{H}_k^T \end{aligned}$$

When applied to the covariance update equation, we obtain

$$P_k = \left(1 - \frac{c_{k-1}}{c_{k-1} + \sigma_k^2}\right) P_{k-1} - 2c_{k-1}^2 \Delta t \dot{H}_k^T H_k \quad (67)$$

where the second term is the time derivative with respect to the time-oscillating matrix elements and the first term is the time derivative with respect to the amplitude decay rate. The result is that  $P$  should obey the following ordinary differential equation (ODE):

$$\dot{P} = \frac{\omega}{\sigma_k^2} (M^T P M - I) \quad (68)$$

where

$$M = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad (69)$$

The effect of time discretization does add an error term into the estimate  $\hat{x}_k$  of  $O(\Delta t)$  because some of the error between the signal and estimated harmonics is erroneously assigned based on the time derivative. The method can be improved by removing the discrete evolution equation for  $P_k$  and replacing it with the exact solution:

$$P_k = \begin{bmatrix} \cos^2(\omega t_k) & \cos(\omega t_k) \sin(\omega t_k) \\ \cos(\omega t_k) \sin(\omega t_k) & \sin^2(\omega t_k) \end{bmatrix} \quad (70)$$

### 3.3. Maximum Amplitude Detection Algorithm

The results of the previous section are applicable to identifying the spectral content of a signal at a prescribed frequency. Efficiently scanning a frequency range necessarily implies using a set of these filters to sample the range of frequencies. Since it is unlikely that the initial set will contain the peak frequency, an adaptive approach must be used if the Kalman filter frequencies are limited to a finite set. In addition to converging to the frequency with maximum energy, the method must also be able to

respond to transients. As will be shown, the Kalman filtering algorithm approximates the DTFT which converges to the time-averaged value of the frequency content. In this section, we will first consider a method for iteratively choosing a Kalman frequency set that finds and tracks the peak frequency. In addition to locating the peak in frequency space, the error associated with its location will also be estimated. Restarting this algorithm after the estimate has sufficiently converged and a peak identified will be discussed. This facility will enable the algorithm to effectively identify transient events. The error reduction rate will determine the maximum restart rate, which, in turn, drives the time resolution of the method.

Now, consider a set of  $2N + 1$  frequencies initially evenly-spaced on the domain  $[\omega^{-N}, \omega^{+N}]$ . The associated KFs are given initial conditions  $a_{n,0} = b_{n,0} = 0$  and  $P_{n,0} = I$  and allowed to observe the signal for a length of time  $T_1$ . After this length of time, a new set of frequencies is determined centered around the maximum frequency of the previous set. Because the Kalman filter algorithm reproduces the DTFT, the width of the lobes in frequency space at time  $t = T_1$  is used to select the initial spread of frequencies of the Kalman filters. Specifically, the critical initial time,  $T_1$ , is when the lobe width  $2\pi/T_1$  treating the signal as a pure harmonic (refer to Section 3.1) is just less than the width of the spread of the primary mode due to intrinsic damping (see Fig. 2 and Ref. [30, Fig. 1]).<sup>j</sup> A theoretical estimate of the fundamental frequency from the beam theory in Section 2 provides the initial center frequency.

Frequency tracking also requires a method for refining the frequency set. The relationship with the DTFT provides a natural framework by letting the width in frequency space match the lobe size at the current time, thus avoiding the nodes where the amplitude goes to zero. We chose to let refinement only occur if the frequency with maximal energy is encountered in the center third of the current frequency set. Otherwise the entire frequency set is shifted in the direction of the maxima. If a minima occurs in the middle third, the frequency set is expanded to promote exploration. In contrast to the initial run,  $T_i$  for  $i > 1$  are reduced because they no longer have to get past the initial large lobe to obtain a good estimate of the frequency content. To obtain a more rapid response rate, we empirically determined that a value of  $T_i = T_1/5$  was appropriate.

A naïve restarting algorithm for stages  $i > 1$  would use the same initial conditions as the first run on the new frequency set. This is disadvantageous for several reasons. Most importantly, knowledge regarding the coefficient amplitude would be lost, as well as the connection with the DTFT used to interpret the results. Maintaining this connection requires interpolating the coefficients  $a_n$  and  $b_n$  to the new set of frequencies. If the exact

<sup>j</sup>Relating peak width to lobe width proved more successful in the trials detailed in the Results section than estimates based on the decay rate of the Kalman filter covariance amplitude.

Kalman filtering algorithm is used, the covariance is also interpolated. Otherwise it is directly set using Eq. (70). While effective for frequencies within the existing band, it is sometimes necessary to expand or shift the frequency range. Extrapolation of the frequency information is used in these cases; however, shifts are limited to not exceed the spacing of the filters in frequency space so as to reduce extrapolation error. By using these methods and restrictions to set the coefficients of the new frequencies, the Kalman filter frequency identification is able to match a long-time DTFT, as will be shown in the next section. Algorithm 1 provides the details of this approach.

**Algorithm 1** Peak Frequency Determination Based on Lobe Decay and Global Restart

```

1: initialize: center frequency:  $\varpi$ , width:  $\Delta\omega$ 
2: period of observation:  $T = T_1$ , counter:  $C = 1$ 
3: while  $T < T_{\max}$  do
4:   initialize:  $a_0 = b_0 = 0$ ,  $P_0 = I$ 
5:   for  $i \in -N, N$  do
6:     set  $\omega_i = \varpi + \frac{i}{N} \Delta\omega$ 
7:     initialize  $KF_i$  using  $a_0, b_0, P_0, \omega_i$ 
8:   end for
9:   while  $t < T$  do
10:    for  $i \in -N, N$  do
11:      update  $KF_i$ 
12:       $t = t + \Delta t$ 
13:    end for
14:  end while
15:  set  $\omega_{\max} = \max_i \omega_i$ 
16:  if  $\omega_{\max} = \omega_{-N}$  then
17:     $\varpi = \omega_{-N} - \Delta\omega$ 
18:  else if  $\omega_{\max} = \omega_{+N}$  then
19:     $\varpi = \omega_{+N} + \Delta\omega$ 
20:  else
21:     $\varpi = \omega_{\max}$ 
22:     $\Delta\omega = \pi/T$ 
23:  end if
24:   $T = T + T_C$ 
25: end while

```

Basing the frequency sampling width on the DTFT lobe size is an appropriate method for estimating a single fixed peak. However, real time detection necessitates identifying transient events. To provide this response time, the entire frequency set is periodically restarted with the initial, broad frequency set and amplitude and covariance coefficients. The rate at which this restart can occur determines the temporal sensitivity of event detection and is driven by the accuracy of the estimate. Improving the accuracy of the peak estimate relative to the length of time considered can be accomplished via a closed loop mechanism, detailed in Algorithm 2. Rather than using the lobe size to provide the sampling width, as in Algorithm 1, the width is adjusted based on the energy content of the set of filters. First, if a maxima occurs in the middle third

of the frequency set, the location and amplitude of the maximum frequency is determined using a parabolic fit over the inner third set of amplitudes (more complex functional forms, e.g., Gaussian and Lorentzian, could be used but at the cost of increased computational time to perform the non-linear regression). Amplitude thresholding is used such that the new frequency range covers all frequencies with some fraction of the energy of the peak using the assumed fitting function. We denote this threshold as  $B$ , so that when  $B = 0$  there is no knowledge as to the location of the peak, and when  $B = 1$ , the peak's location is completely determined. Convergence is facilitated by increasing  $B$  after every restart in which the peak frequency occurred in the middle third according to:

$$B = 1 - \frac{\Delta B}{C}. \quad (71)$$

The parameter  $\Delta B$  regulates how quickly  $B$  approaches unity, and  $C$  is the count of times in which the peak frequency occurred in the middle third, i.e., a maximum was found within the frequency set. By using this closed-loop resetting algorithm (Algorithm 2), the convergence of the peak frequency is significantly enhanced, as will be demonstrated in the Results section, thereby increasing the global restart rate which determines the overall sensitivity of the method.

**Algorithm 2** Peak Frequency Determination Based on Sample Variance and Local Restart

```

1: initialize:  $\varpi, \Delta\omega, \Delta T_{L,0}, \Delta T_L, \Delta T_G, \Delta B, C = 1, t = t_0$ ,  

    $c = c_0, a_i = b_i = 0$ ,
2:  $T = t_0 + \Delta T_{L,0}$ ,  $B = 1 - \Delta B$ 
3: while  $T < T_{\max}$  do
4:   for  $i \in -N, N$  do
5:     set  $\omega_i = \varpi + \frac{i}{N} \Delta\omega$ 
6:     set  $P_i$  using Eq. (66),  $\omega_i, c$ , and  $t$ 
7:     initialize  $KF_i$  using  $a_i, b_i, P_i, \omega_i$ 
8:   end for
9:   while  $t < T$  do
10:    for  $i \in -N, N$  do
11:      update  $KF_i$ 
12:    end for
13:    end while
14:     $A_{-1} = \text{ave}(|a_i|^2 + |b_i|^2)$  for  $i \in -N, -N/3$ 
15:     $A_0 = \text{ave}(|a_i|^2 + |b_i|^2)$  for  $i \in -N/3, N/3$ 
16:     $A_{+1} = \text{ave}(|a_i|^2 + |b_i|^2)$  for  $i \in N/3, N$ 
17:    if  $\max_{A_{-1}, A_0, A_{+1}} = A_0$  then
18:       $\Delta\omega = 2\Delta\omega$ 
19:       $C = \max(C - 1, 1)$ 
20:    else if  $\text{argmax}_i[|a_i|^2 + |b_i|^2] = -N$  then
21:       $\varpi = \varpi - \Delta\omega$ 
22:    else if  $\text{argmax}_i[|a_i|^2 + |b_i|^2] = +N$  then
23:       $\varpi = \varpi + \Delta\omega$ 
24:    else
25:       $\varpi = \text{argmax}_{\omega} g(\omega)$  where  $g(\omega)$  is the parabolic  

      fit using  $(\omega_{\max-1}, \omega_{\max}, \omega_{\max+1})$ 

```

```

26:  $\omega_{lo} = \max_{j < \max} \omega_j$  such that  $|a_j|^2 + |b_j|^2 <$ 
 $B \max_i (|a_i|^2 + |b_i|^2)$ 
27:  $\omega_{hi} = \min_{j > \max} \omega_j$  such that  $|a_j|^2 + |b_j|^2 <$ 
 $B \max_i (|a_i|^2 + |b_i|^2)$ 
28: if  $lo = -N$  then
29:  $\varpi = \varpi - \Delta\omega$ 
30: else if  $hi = +N$  then
31:  $\varpi = \varpi + \Delta\omega$ 
32: else
33:  $\Delta\omega = \max(\omega_{\max} - \omega_{lo}, \omega_{hi} - \omega_{\max})$ 
34:  $C = C + 1, B = 1 - \Delta B/C$ 
35: end if
36: end if
37: if  $T - t_0 > \Delta T_G$  then
38:  $t_0 = T, T = t_0 + \Delta T_{L,0}, \Delta\omega = \Delta\omega_0, B = 1 - \Delta B,$ 
 $C = 1, c = c_0, a_i = b_i = 0$ 
39: else
40: for  $i \in -N, N$  do
41: interpolate  $a_n, b_n$  using polar coordinates
42: end for
43:  $T = T + \Delta T_L$ 
44: end if
45: end while

```

## 4. RESULTS

In this section, we demonstrate results of applying the algorithms developed in the previous section to a sequence of models of the CNT sensor increasing complexity culminating in a MD simulation of a CNT with a tip mass and the thermal noise intrinsic to the CNT and substrate. Before employing the KF algorithm on the MD data, we validate it on the computationally less expensive and parametrically simpler noisy sine and Langevin (see Section 2.2) model systems. Both models are simple enough to lend themselves to analysis, yet contain the basic features of our target system and, hence, are good surrogates for thermally-excited CNT tip dynamics. In particular, Langevin dynamics have more features particular to the thermalized CNT but requires more sophisticated analysis and therefore is a stepping-stone to the full CNT dynamics.

For these model systems, we show:

- (a) KF spectral density correlates well with the DTFT and FFT of the signal,
- (b) rate of convergence of the (incrementally updated) KFs is comparable or better than the DTFT, and
- (c) the KF is robust with respect to noise.

The theory developed in Sections 2 and 3 is used to provide expectations of the system response and appropriate model parameters. In addition, we use fine-grained DTFT frequency scans and ensemble averages as direct estimates of the stochastic response these surrogates for the CNT tip dynamics.

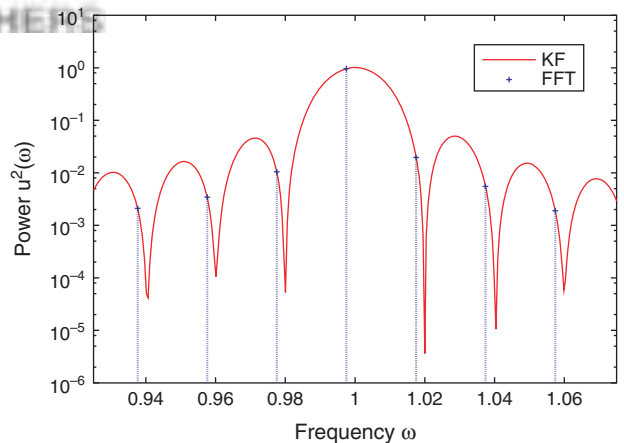
### 4.1. Noisy Harmonic

For sample interval  $\Delta t$  and period  $P = 2\pi/\omega_0$ , a normalized noisy sinusoid is simply

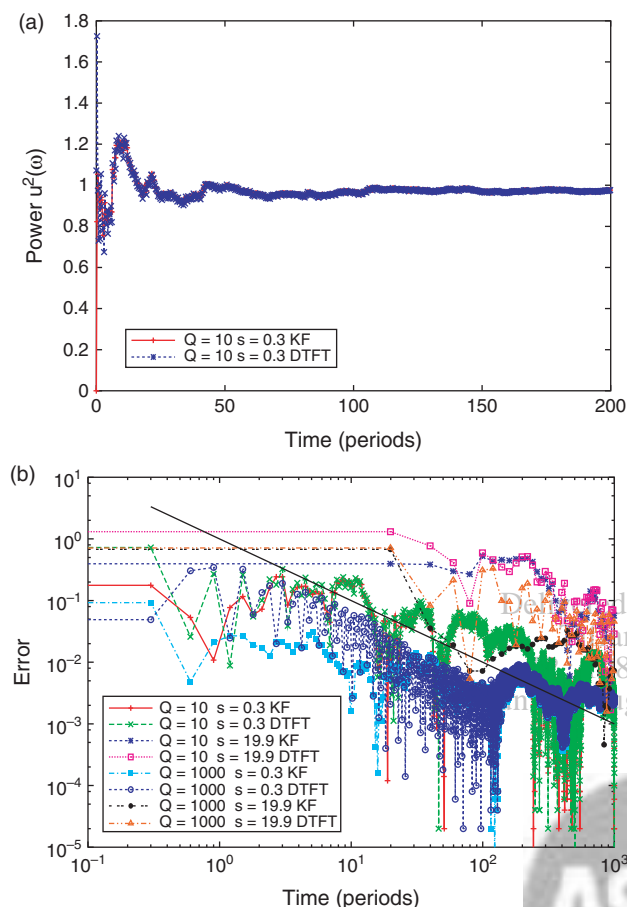
$$u_{t_i} = \sin(\omega_0 \Delta t i) + \sqrt{\frac{\Delta t}{Q}} w_i \quad (72)$$

where we chose  $\Delta t = P/100$ ,  $\omega_0 = 1$  and  $w_i$  to be sampled from the zero mean, unit variance normal distribution  $\mathcal{N}(0, 1)$ . The Kalman filter measurement variance is  $\sigma^2 = \Delta t/Q$  consistent with the process noise. Clearly, for this signal,  $c_0 = \langle u^2 \rangle \equiv \mathcal{A}_u(0) = \frac{1}{2}$  and  $\langle u^2(\omega_0) \rangle = 1 + \Delta t/Q \approx 1$  since  $\langle w_i \sin(\omega_0 t) \rangle = 0$ , and  $\Delta t \ll 1$  and  $Q \gg 1$ . Other initial conditions are specified in Section 3.3. This model follows the expectations of KF for form of model and noise. As the goal of this section is to examine the Kalman filtering frequency estimation only, algorithms to refine the frequency are not used in favor of a large set of filters which sample a range of frequencies.

Figure 5 shows that the lobes characteristic of the sinc function (discussed in Section 3) are recovered by a frequency partition of KFs. This data is nearly identical to that recovered by a corresponding DTFT, the differences due to how the KF handles noise will be illustrated in detail in next section. The FFT data is also shown, which demonstrates how the devised KF (and the DTFT) can interpolate the spectrum at arbitrary frequencies with respect to the FFT. (The FFT has a frequency spacing determined by the length of the signal as discussed in Section 3.1.) We define the error in the amplitude as  $||\mathbf{x}_\omega||^2 - u^2(\omega)|^2$ , where  $\mathbf{x}_\omega = [a, b]$  is the estimate from a KF at  $\omega$ , and  $u^2(\omega)$  is the expectation of the signal at the given frequency  $\omega$ . Figure 6 shows the evolution of the power estimate  $||\mathbf{x}_\omega||^2$  and the convergence with respect



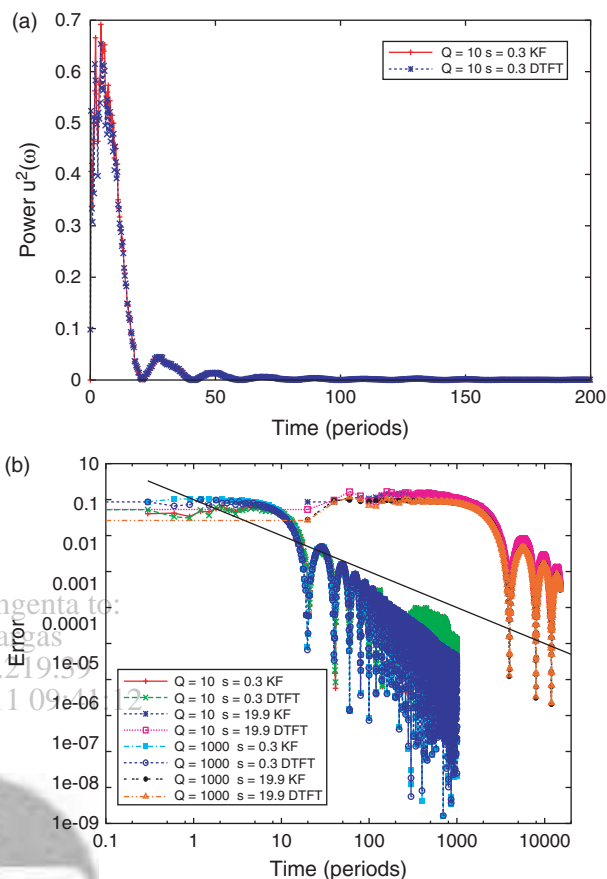
**Fig. 5.** Near peak spectral response for a noisy sine ( $Q = 100$ ). The lobes characteristic of the DTFT are clearly visible in the KF. Also the KF clearly agrees with the FFT, which here is sampled off frequency for illustration of spectra content away from the max. For a FFT sampled on the period of the signal, the off frequency spectral estimates sit at the zeros between the lobes.



**Fig. 6.** Evolution (a) and convergence (b) of the KF and DTFT estimates of a noisy sine spectrum at the peak frequency  $\omega = \omega_0$ . The black trend line in (b) has exponent  $-1$  and the sampling frequency  $s$  is relative to the expected period  $P$  of the sine.

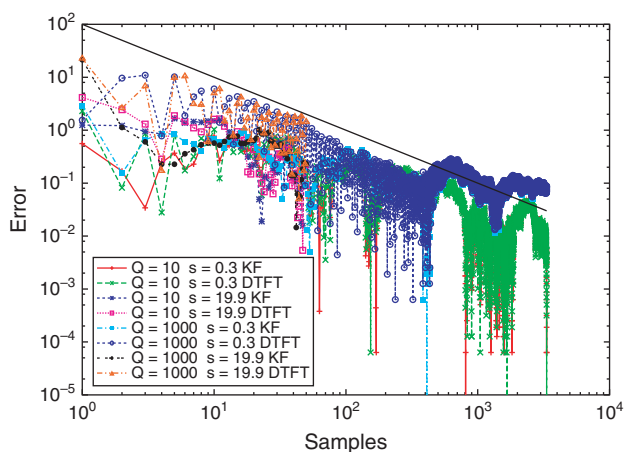
to the expectation for the on-peak frequency  $\omega = \omega_0$  for both low and high  $Q$  systems, as well as sampling the signal well below and well above the period  $P$ . We compare the convergence to evolution of DTFT at same frequencies which on-average have worse errors. Figure 7 shows evolution and convergence to zero amplitude for an off-peak frequency  $\omega = 0.95\omega_0$ . It also shows the typical convergence is faster than the on-peak case. (In fact, the decay rates for off-peak frequencies depend strongly on distance from the peak frequency with frequencies farther from the peak decaying faster consistent with decay rate of the peaks of the sinc function.) Also apparent is that both the KF and the DTFT show artifacts in their convergence that are directly related to the lobes in the spectrum passing through the given frequency as  $T$  increases.

Lastly, from the approximate universal relation in Figure 8 we see that the convergence of the error in the amplitude  $\|\mathbf{x}_\omega\|^2 - u^2(\omega)\|^2 \sim N^{-1} \sim T^{-1}$  with the leading coefficients of the decay rates only marginally dependent on noise and sampling frequency. Here,  $N$  is number of samples and  $T$  is the total observation time. Hence, we conclude that the efficiency in determining the frequency



**Fig. 7.** Evolution (a) and convergence (b) of the KF and DTFT estimates of a noisy sine spectrum off the peak frequency at  $\omega = 0.95\omega_0$ . The black trend line in (b) has exponent  $-1$  and the sampling frequency  $s$  is relative to the expected period  $P$  of the sine.

content is relatively insensitive to sampling period. It is also apparent that in some cases that there is some propensity for the KF to saturate and its error estimate to level off. We believe this is due to the fact that as  $c_i \rightarrow 0$  the



**Fig. 8.** Convergence of the on-peak estimate with respect to the number of samples showing sampling efficiency. The black trend line has exponent  $-1$ .



new samples contribute very little to the estimate. If, in fact this is the cause of the saturation, the DTFT should be relatively immune to it.

## 4.2. Langevin Dynamics

A more complex model system that arguably better represents the CNT's response is the non-dimensional harmonic Langevin developed in Section 2.2. It differs from the noisy sine in that brown instead of white noise is observed in the signal  $u_{t_i}$ , i.e., the noise has a non-uniform spectral representation, as the white noise in the input is shaped by the system response. It also differs in that the noise is not additive, as assumed in the KF, although as discussed in Section 3.2, as long as the noise manifested in the observations is approximately white, the KF method is appropriate and effective. Following<sup>43,44</sup> we integrate (34) via<sup>k</sup>

$$\begin{aligned} v_{i+1/2} &= v_i + \frac{1}{2} \Delta v_i \\ u_{i+1} &= u_i + v_{i+1/2} \Delta t \\ \Delta v_{i+1} &= -u_{i+1} \Delta t - \frac{1}{Q} (v_{i+1/2} \Delta t - \sqrt{Q \Delta t} w_{i+1}) \\ v_{i+1} &= v_{i+1/2} + \frac{1}{2} \Delta v_{i+1} \end{aligned} \quad (73)$$

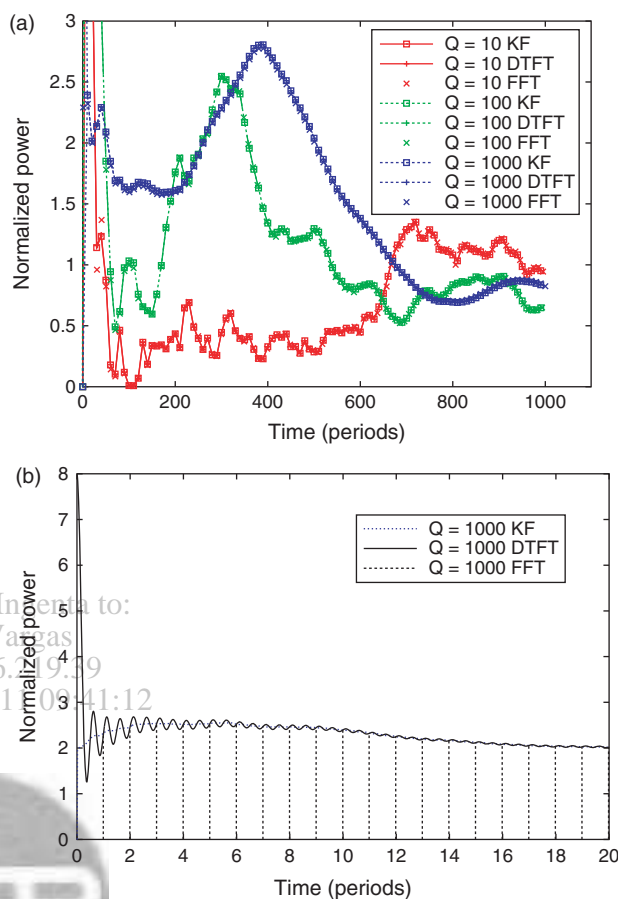
using a time-step  $\Delta t = P/100$ . The period  $P \approx 2\pi$  for high enough  $Q$  since  $\omega_0 = 1$  is implicit in Eq. (73), and in (34) which it approximates. As stated in Section 2.2, there are two significant parameters:  $Q$ , and  $\Delta t \omega_0$  which in turn determine the variance of the non-dimensional white noise  $\sigma^2 = \Delta t/Q$  and the maximum intensity  $\mathcal{F}(\omega_0) = Q/\pi$ .

In Figure 9 we compare the KF response to sampling the signal  $u_{t_i}$  at an interval  $s = 0.9P$  at fixed frequency given by the frequency at maximum amplitude of the FFT at  $10^4$  samples. Figure 9(a) shows that the KF, DTFT and FFT estimates are all in agreement in the long run over a range of quality factors. Figure 9(b) clearly shows that the KF estimate has much reduced variance at short time compared to the DTFT due to its explicit handling of noise.

Now using the parallel KF Algorithm 1 we estimate the evolution of the peak power. Figure 10 shows the trajectory of the peak amplitude for various instantiations of the harmonic Langevin process. The background color contours are calculated from an exhaustive scan of the process normalized to be in the range  $[0, 1]$  at each time slice. Clearly the tracking of the peak power is robust despite fluctuations and discontinuities in the underlying power spectrum in some of the ensemble.<sup>l</sup> Figure 11a shows the ensemble

<sup>k</sup>The numerical approximation to the random force will be band-limited at frequency  $2\pi/\Delta t$  and correlated over  $\Delta t$ , so Eq. (25) holds only approximately.

<sup>l</sup>Note that we use a random number generator to construct the sequence  $w_i$  but perhaps sine series based on the Karhunen-Loève theorem would have been more appropriate for CNT higher modes. Empirically, the results in Section 4.3 display distinctly more continuous power spectra over time for the few instances we calculated.



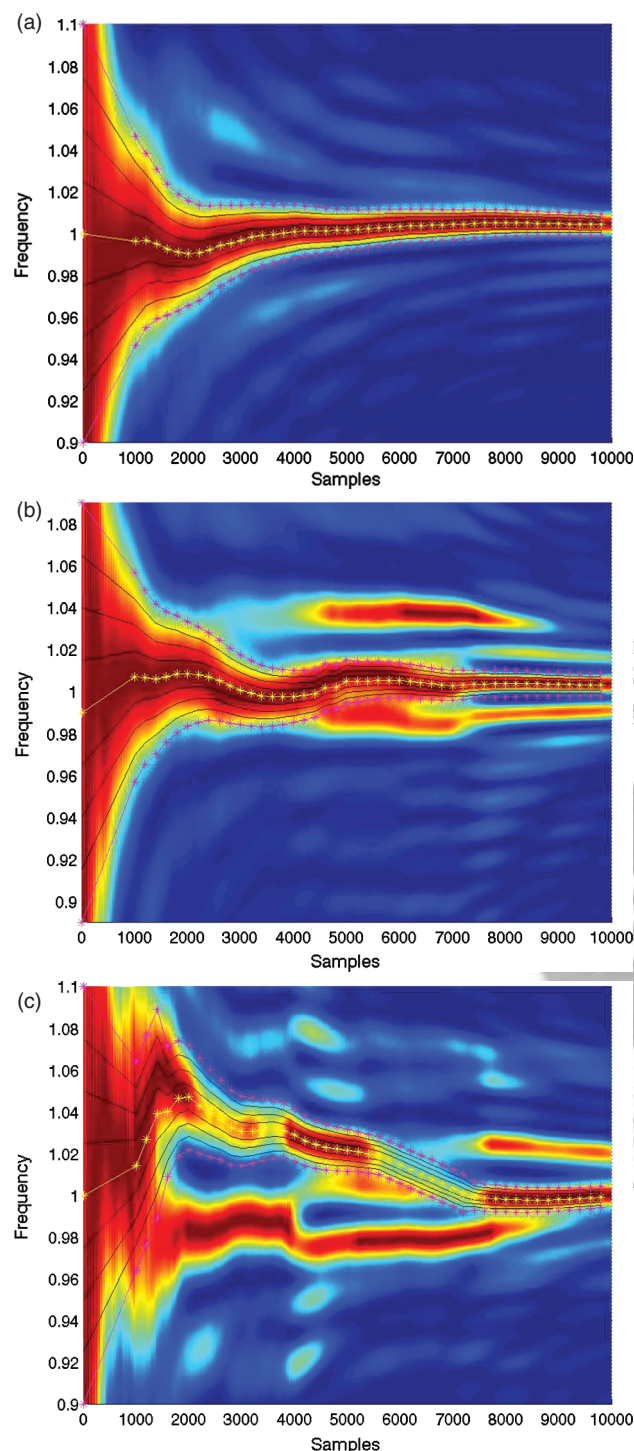
**Fig. 9.** Comparison of KF, DTFT and FFT for the Langevin signal. From (b) it is clear that KF has significantly less noise than DTFT at early times. Both appear to interpolate FFT values. The estimated power has been normalized by  $Q/\pi$ .

of long-time estimates from 100 separate instantiations of the Langevin process. Notice that frequency distribution is peaked whereas the amplitude distribution is widely spread about the expectation of 1. We see this as more evidence that frequency-based detection of attached mass is superior to amplitude-based. Figure 11(b) shows the evolution of the frequency estimate over the ensemble and that the error estimated from the variance of the frequency ensemble appears to asymptote to a constant, consistent with the spread about the peak frequency caused by internal damping.

## 4.3. Molecular Dynamics

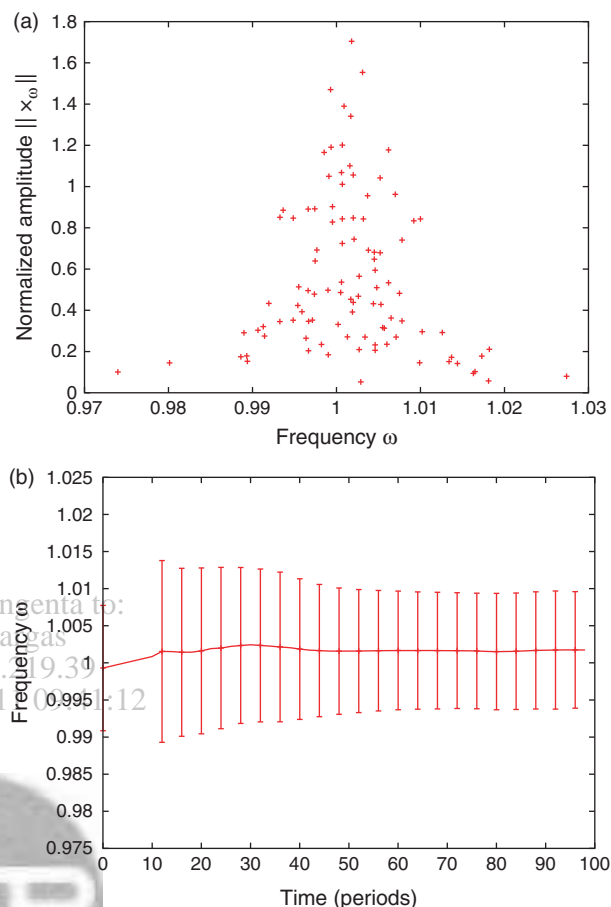
Molecular dynamics (MD) is typically employed to model the dynamics of molecular systems and is an ideal tool for representing phonon modes and thermal noise. MD gives the stable lattice structure, computes the correct anharmonic phonon interactions, and the finite number of eigenvalues of the thermal spectra. It is founded on Newton's law

$$m_\alpha \ddot{\mathbf{x}}_\alpha = \mathbf{f}_\alpha \quad (74)$$



**Fig. 10.** Path of the adaptive set of KFs for various instances of the Langevin model generated by different realizations of the noise process overlaid on an exhaustive sweep by the DTFT. Background colors denote the relative value of any point to the peak taken at each time slice. Red indicates a value of 1 (identical to peak) while blue denotes zero.

for the atoms (ion cores) with positions  $\mathbf{x}_\alpha$  and masses  $m_\alpha$  together with an interatomic force  $\mathbf{f}_\alpha = \partial\Phi/\partial\mathbf{x}_\alpha$  derived from an empirical potential  $\Phi$ . Here  $\alpha$  is an index over all the atoms.



**Fig. 11.** Ensemble of the peak amplitude and frequency estimates at time 100P ( $10^4$  samples) (a) and evolution of the frequency estimate (b).

As mentioned earlier, the configuration we employ is illustrated in Figure 1 where a single wall (10, 10) CNT, modeled with the AI-REBO potential,<sup>45,46,m</sup> is cantilevered 17.8 nm from a substrate formed of Au atoms following an EAM potential.<sup>49</sup> The interaction between the CNT and the Au substrate was modelled with a simple Lennard-Jones potential (with parameters  $\sigma = 1.0$  Å and  $\epsilon = 1.0$  eV)<sup>n</sup> appropriate for surface interactions like van der Waals attraction. As opposed to placing the CNT directly under a thermostat's control or omitting temperature control entirely, we choose to explicitly represent the substrate in order to best simulate the thermal fluctuations of the CNT accurately in real time. We embed the CNT in the substrate which is, itself, under the control of a Nosé-Hoover (NH) thermostat<sup>50,51</sup> while the CNT undergoes Newtonian dynamics according to (74). The contact between the substrate composed of 31648 Au atoms and

<sup>m</sup>The AI-REBO potential was designed to represent hydrocarbons, CNTs and non-bulk carbon phases as opposed to Tersoff potential<sup>47,48</sup> which was tuned to bulk states.

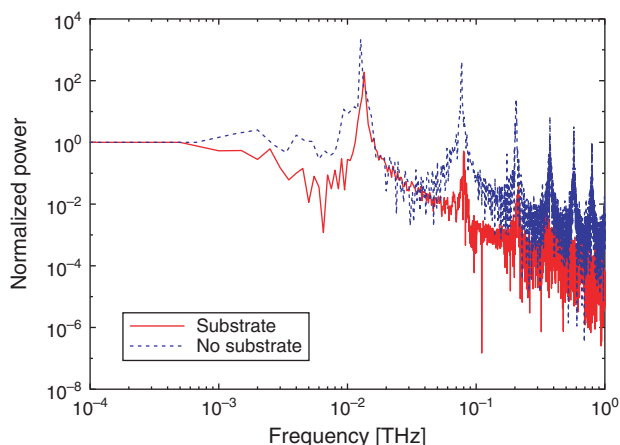
<sup>n</sup>We found that the resonance spectrum of the CNT's tip was largely insensitive to the particular values of the CNT-Au interaction over a wide range of values.

680 C atoms that compose the embedded part of the CNT (2900 C atoms are in the cantilevered section) provides temperature regulation of the CNT. The NH thermostat allows the substrate to sample the canonical, constant temperature ensemble as if it were in contact with a much larger, implicit thermal reservoir. We performed all simulations at room temperature  $\theta = 300$  K and thermalize the system by selecting initial velocities from the Boltzmann distribution (separately for the CNT and substrate) and then allowing the system to relax under thermostat control. In addition, we fixed the 3 outer layers of substrate, except on side of cube where the CNT is embedded, to provide a statically-determinant base for the CNT vibrations.

We map the atomic displacement,  $\mathbf{u}_\alpha = \mathbf{x}_\alpha - \mathbf{X}_\alpha$  where  $\mathbf{X}_\alpha$  are positions in an undeformed reference configuration, to a continuous, transverse deflection  $\mathbf{u}(x, t)$  using cross-sectional averages and interpolation. Even with the constraints on the substrate, the assumed boundary conditions  $\mathbf{u} = \mathbf{0}$  and  $\mathbf{u}' = \mathbf{0}$  only hold on average at the embedded end.<sup>o</sup> Assuming a fixed boundary condition at the embedded end and using the value  $EI = 0.40806 \mu\text{N}\text{\AA}^2$  from Ref. [30] (which employed the Tersoff potential and a rigidly fixed, zero temperature boundary condition instead of an explicit substrate), we estimate the fundamental frequency of the CNT  $\omega_1 = 0.0197$  THz and the period  $P = 0.0507$  ns. In Figure 2, we observe a discrepancy  $\omega_1 \approx 0.0143$  THz from this estimate that was highly accurate for the configuration given in Ref. [30]. The difference is most likely due to the softer, thermal attachment in the present case or the different potential has different elastic properties. To test these hypotheses, we ran a configuration with a rigid end instead of a substrate. Figure 12 shows that the fundamental resonance occurs at essentially the same frequency as with the substrate thus giving credence to the hypothesis that the Tersoff and AI-REBO potentials have slightly different elastic properties. This figure also shows that the explicit substrate tends to damp the higher resonances preferentially. See also Figure 2 for an illustration of the multiple higher resonances that compose the thermal/noise bath for the fundamental mode. On a related note, a relaxation time to disturbances is  $\tau \sim m/\gamma = Q(P/2\pi)$  based on the damping due to interactions of the fundamental with the higher modes implicit in the Langevin model. So, given a quality factor on the order of  $Q \in (100, 1000)$ , this scaling leads to relaxation times on the order of 1–10 ns. Consequently we were very careful to prepare and thermalize the CNTs before taking data, but this effect was difficult to entirely negate.

To add the effect of an attached molecule, we increased the mass of the last two rings of atoms uniformly, as alluded to in the Introduction. So adding  $20 m_C \alpha$ , where  $m_C$  is the mass of a carbon atom, lead to a mass ratio  $m/M = (20/2900)\alpha = 0.0069\alpha$ .

<sup>o</sup>We also assume that the correlations  $\langle \mathbf{u}'' \cdot \mathbf{u}' \rangle$  and  $\langle \mathbf{u}''' \cdot \mathbf{u} \rangle$  relevant to Eq. (12) are small.

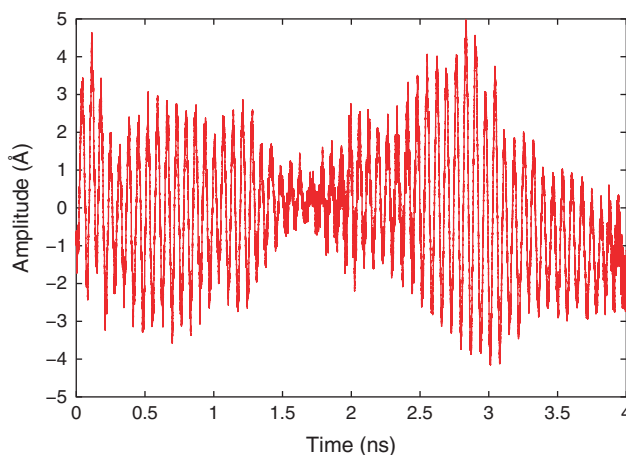


**Fig. 12.** Comparison of CNTs with and without a substrate. In the absence of an explicit substrate the CNT's embedded end is rigidly fixed.

#### 4.3.1. Minimum Sampling Time

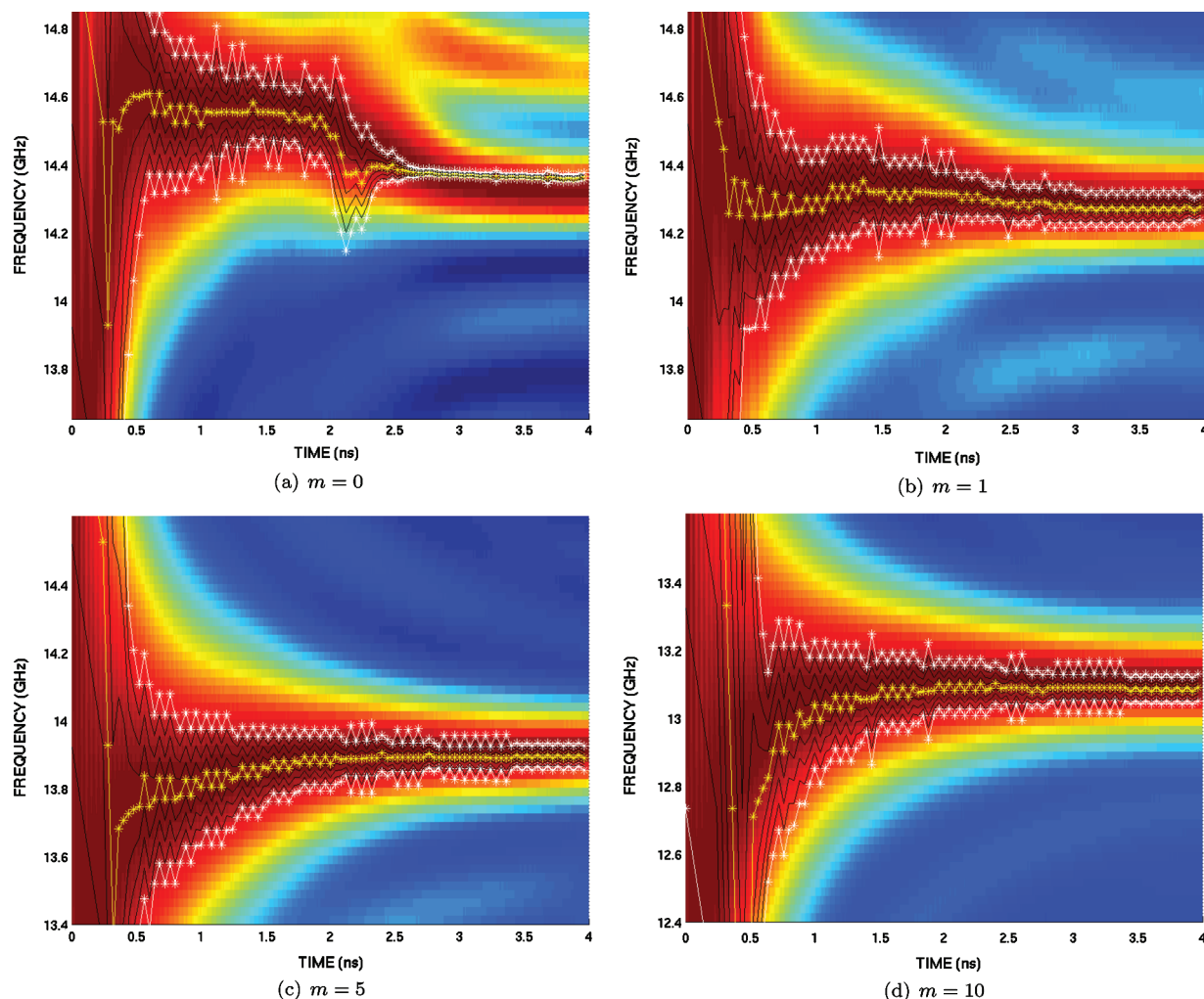
To determine the shift in the fundamental frequency of the CNT for a sequence of attached masses we employ the more sophisticated, closed-loop variant of the parallel KF, namely Algorithm 2. The system is initialized in this case so as to mimic the attachment of a molecule at the tip for a long time, i.e., sufficiently long for the system to relax to equilibrium. Figure 13 shows the tip trajectory for the CNT without a tip mass that we used as a baseline. This trajectory displays considerable fluctuations that are also apparent in the background scan shown in Figure 14(a). The sequence of frequency spectra shown in Figure 14 display a trend toward cleaner spectra as the tip mass is increased (which translates to better signal-to-noise ratio for this freely vibrating system). This is expected, as theory predicts a slower vibration with added mass together with an increased amplitude of the fundamental mode.

Figure 14 as a whole also shows that the proposed KF is very effective at tracking the peak and that Algorithm 2 achieves confidence bounds on the order of 0.1%, which



**Fig. 13.** Tip trajectory for  $m/M = 0$ .





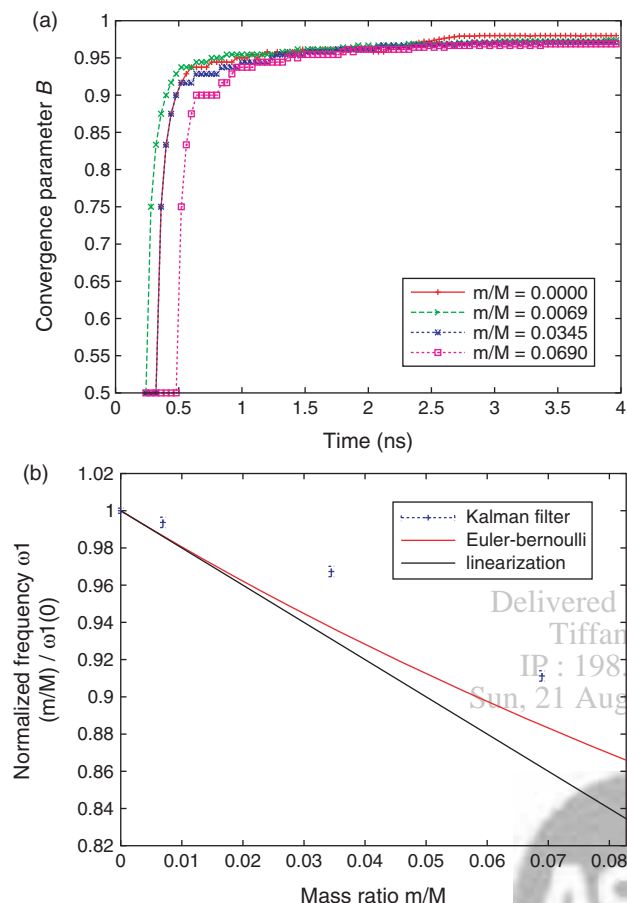
**Fig. 14.** Kalman tracking of the peak amplitude from thermally vibrating CNT with an attached mass. Note the vertical axis in each plot has been selected to focus on the peak frequency to display the details of the tracking algorithm.

is dramatically better than the simpler Algorithm 1. The sequence of frequency spectra shown in Figure 14 shows that the convergence for all cases given is comparable. However, if the fact that the trajectories with larger masses took longer to start tracking, presumably due to the fact that the initial guess for the frequency based on an unloaded beam is further from the truth, is taken into account, the convergence in the frequency estimates shown in Figure 15(a) displays a dependence on the magnitude of the attached mass corresponding to the observations made on Figure 14, i.e., the larger the mass the cleaner the signal and the more rapid the convergence. Figures 14 and 15(a) illustrate the trade-off between certainty in the frequency shift and the period of observation. Even at 4 ns the error is approximately 0.2% given an ideal detector. Also of interest is the deviation of the data from both the full beam theory and the often-employed linearized model shown in Figure 15(b). We speculate that this is due to the softer, noisier connection provided by the thermalized substrate, although it is possible that the discrepancy is simply

noise that ensemble averaging would remove. Given that the deviation is systematic it makes the latter conjecture unlikely. Also of note is the discrepancy between the linearized theory and the full model even in the low percents of  $m/M$ , a fact which has practical impact on the accuracy of mass sensors in and of itself.

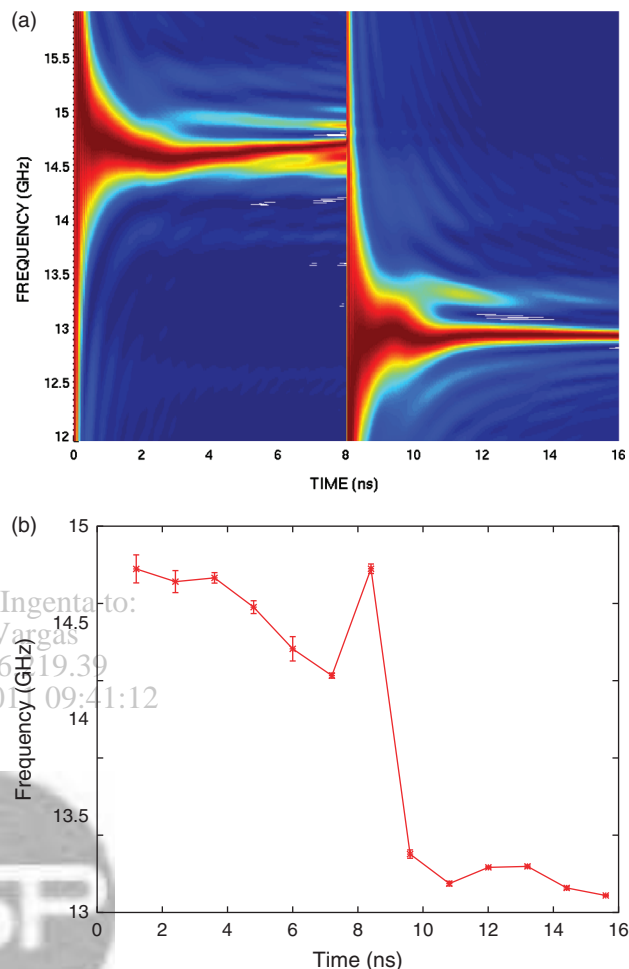
#### 4.3.2. Finite Residence Time

In this example we simulate the rudiments of a molecule that becomes attached and stays attached for a finite amount of time. This time may not be sufficient for the system to attain equilibrium. We simulate this process by simply changing the mass of the atoms at the tip of the CNT so that  $m/M$  jumps from 0 to 0.0690 at a specified time. The background scans in Figure 16(a) clearly shows the change in spectral content and a shift in frequency comparable to that calculated in the previous section for  $m/M = 0.0690$ . To determine the time of attachment, the Kalman filters are periodically restarted completely using



**Fig. 15.** (a) convergence of each of the four cases  $m/M = 0, 0.0069, 0.0345, 0.0690$  based on the convergence parameter  $B$  in Eq. (71) and (b) relative frequency shifts from KF estimates compared to both full beam theory and a linearized model.

only the center frequency from the previous run as input. These frequency estimates also demonstrate a large shift when the mass attaches, as shown in Figure 16(b). Note the full frequency scan was only restarted once over the duration considered, i.e., when the mass attaches mid-way through, while the Kalman tracking algorithm resets approximately every nanosecond. This resetting frequency determines the temporal response time. The error in the mass estimate, on the other hand, is directly related to the estimated frequency shift, as in the last section. Also apparent from Figure 16 is that the correlation time for disturbances of the fundamental mode are considerably shorter than the estimate  $\tau \sim Q(P/2\pi)$  would lead us to believe. It is possible that the short-wavelength thermal vibrations have different relaxation times than the fundamental, and the cascade from the fundamental to the higher modes of the thermal bath and vice versa is likely much slower and on the timescale of the estimated relaxation time. Since we are only interested in the response of the fundamental mode, we are not hindered by the conjectured slow relaxation between modes. Figure 16 demonstrates the inner/outer loop mode to reset the trajectory tracking



**Fig. 16.** Kalman tracking the change in the fundamental CNT tip frequency as a mass attaches. The global scan (top) applies the Kalman filtering algorithm over a large frequency range at the initial time and again at the attachment time. Frequency tracking is performed by periodically restarting the Kalman ensemble and using the estimate for the peak at the end of an ensemble (bottom).

is effective in estimating the time a mass attaches to the tip to the resolution of approximately 1 ns.

## 5. DISCUSSION

We explored the limits of mass detection via simulation that accounted for the thermal effects associated with anharmonic phonon interaction in the CNT and interactions with a finite temperature substrate, albeit idealized in other aspects. The development of the proposed filtering algorithms was facilitated by theoretical developments that provided a fundamental understanding of the noise characteristics of a freely vibrating CNT. This aspect of the work included: (a) determining dependence of resonance frequencies on added tip mass through the characteristic equation, Eq. (7), (b) an extension of the Treacy<sup>8</sup> equilibrium theory to the case of additional mass attached

to the tip, Eq. (15), to determine the amplitude sensitivity to added mass, and (c) a reduced-order Langevin model that provided much of the characteristic dynamics in a simplified form, as well the peak intensity of the anharmonically damped fundamental resonance, Eq. (30), and a non-dimensional version of the relevant parameters, Eq. (34). Guided by this theory and with reference to basic spectral analysis, we developed a Kalman filter, specifically Algorithm 2, customized to do real-time/on-the-fly estimation of the necessary spectral information to detect shifts in the primary resonant frequency. This novel Kalman filter for spectral analysis proved to be not only efficient due to its incremental nature but also very effective and noise-tolerant in tracking the trajectory of the peak amplitude through frequency space. With our algorithm and an ideal detector we demonstrated that mass discrimination on the order of fractions of a percent can be achieved in nanoseconds. Moreover, our Kalman filter can effectively undersample the dynamics to accommodate the frequency limits of existing detectors and still potentially achieve orders of magnitude improvement in the response time of resonator-based mass detectors.

While this work concentrated on detecting attached masses by tracking changes in the fundamental frequency of thermal vibrations, many other potential uses for these methods exist. For example, the proposed algorithm can be easily modified to identify changes for multiple, well-separated resonant peaks of the CNT (with a corresponding added burden on the bandwidth of the detector). In future work we intend to explore the dynamics of a CNT driven by an external electric field, as commonly used in many nanoscale mass detectors, that is also under the influence of thermal noise like the freely vibrating CNT analyzed in this work. Also, we would like to investigate the interaction of the vibrating CNT and the ambient atmosphere including more of the details of the dynamics and kinetics of molecule attachment and release. In addition, we intend to develop a noise model for a sufficiently realistic detector based on one of the existing technologies as a prelude to implementing our algorithms in a real device. In this aspect, the proposed parallel Kalman filter has a straightforward graphics processing unit (GPU) implementation.

**Acknowledgments:** We would like to acknowledge the support of the Laboratory Directed Research and Development program at Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-ACO4-94AL85000.

## References

1. H. G. Craighead, *Science*, 290, 1532 (2000).
2. K. L. Ekinci and M. L. Roukes, *Rev. Scientific Instruments* 76, 061101 (2005).
3. K. Jensen, K. Kim, and A. Zettl, *Nat. Nano.* 3, 533 (2008).
4. B. Ilic, H. G. Craighead, S. Krylov, W. Senaratne, C. Ober, and P. Neuzil, *J. Appl. Phys.* 95, 3694 (2004).
5. K. L. Ekinci, X. M. H. Huang, and M. L. Roukes, *Appl. Phys. Lett.* 84, 4469 (2004).
6. K. L. Ekinci, Y. T. Yang, and M. L. Roukes, *J. Appl. Phys.* 95, 2682 (2004).
7. Y. T. Yang, C. Callegari, X. L. Feng, K. L. Ekinci, and M. L. Roukes, *Nano Lett.* 6, 583 (2006).
8. M. M. J. Treacy, T. W. Ebbesen, and J. M. Gibson, *Nature* 381, 678 (1996).
9. P. Poncharal, Z. L. Wang, D. Ugarte, and W. A. De Heer, *Science* 283, 1513 (1999).
10. A. N. Cleland and M. L. Roukes, *Appl. Phys. Lett.* 69, 2653 (1996).
11. A. N. Cleland and M. L. Roukes, *Sensors and Actuators A-Physical* 72, 256 (1999).
12. V. Sazonova, Y. Yaish, H. Ustunel, D. Roundy, T. A. Arias, and P. L. McEuen, *Nature* 431, 284 (2004).
13. H. B. Peng, C. W. Chang, S. Aloni, T. D. Yuzvinsky, and A. Zettl, *Phys. Rev. Lett.* 97, 087203 (2006).
14. X. L. Feng, R. He, P. Yang, and M. L. Roukes, *Nano Lett.* 7, 1953 (2007).
15. H. J. Mamin and D. Rugar, *Appl. Phys. Lett.* 79, 3358 (2001).
16. M. D. Lahaye, O. Buu, B. Camarota, and K. C. Schwab, *Science*, 304, 74 (2004).
17. K. Jensen, J. Weldon, H. Garcia, and A. Zettl, *Nano Lett.* 8, 374 (2008).
18. P. Mohanty, D. A. Harrington, K. L. Ekinci, Y. T. Yang, M. J. Murphy, and M. L. Roukes, *Phys. Rev. B* 66, 085416 (2002).
19. K. Jensen, H. Peng, and A. Zettl, Limits of nanomechanical resonators, *Proceedings of the 2006 International Conference Nanoscience Nanotechnology*, IEEE, Piscataway, NJ, USA (2006).
20. A. N. Cleland and M. L. Roukes, *Nature*, 392, 160 (1998).
21. R. G. Knobel and A. N. Cleland, *Nature* 424, 291 (2003).
22. A. N. Cleland and M. L. Roukes, *J. Appl. Phys.* 92, 2758 (2002).
23. R. Ciocan, J. Gaillard, M. J. Skove, and A. M. Rao, *Nano Lett.* 5, 2389 (2005).
24. R. C. Batra and A. Sears, *Int. J. Sol. Struct.* 44, 7577 (2007).
25. M. Mitra and S. Gopalakrishnan, *J. Appl. Phys.* 101, 114320 (2007).
26. M. B. Viani, T. E. Schaffer, A. Chand, M. Rief, H. E. Gaub, and P. K. Hansma, *J. Appl. Phys.* 86, 2258 (1999).
27. A. A. Maradudin and A. E. Fein, *Phys. Rev.* 128, 2589 (1962).
28. A. A. Maradudin, A. E. Fein, and G. H. Vineyard, *Phys. Stat. Solidi.* 2, 1479 (1962).
29. H. Michel and M. Wagner, *Phys. Stat. Solidi B* 85, 195 (1978).
30. E. H. Feng and R. E. Jones, *Phys. Rev. B* 81, 125436 (2010).
31. P. A. A. Laura, M. J. Maurizi, and J. L. Pombo, *J. Sound and Vibration* 41, 397 (1975).
32. M. Gürgöze, *J. Sound and Vibration* 282, 538 (2005).
33. P. A. A. Laura, J. L. Pombo, and E. A. Susemihl, *J. Sound and Vibration* 37, 161 (1974).
34. K. Huang, *Statistical Mechanics*, Wiley, Hoboken, NJ, USA (1987).
35. A. Krishnan, E. Dujardin, T. W. Ebbesen, P. N. Yianilos, and M. M. J. Treacy, *Phys. Rev. B* 58, 14013 (1998).
36. B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th edn., Springer, Berlin (2003).
37. R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics I: Nonequilibrium Statistical Mechanics*, 2nd edn., Springer, Berlin (1991).
38. G. L. Bretthorst, *Journal of Magnetic Resonance* 88, 571 (1990).
39. J. Gronczynski, *Signal Processing* 87, 1003 (2007).
40. A. H. Sayed and T. Kailath, *IEEE Signal Processing Magazine* 11, 18 (1994).
41. A. Holland and M. Aboy, *Medical and Biological Engineering and Computation* 47, 697 (2009).
42. R. E. Kalman, *Journal of Basic Engineering* 82, 35 (1960).



43. A. Brünger, C. L. Brooks, and M. Karplus, *Chem. Phys. Lett.* 105, 495 (1984).
44. R. W. Pastor, B. R. Brooks, and A. Szabo, *Mol. Phys.* 65, 1409 (1988).
45. S. J. Stuart, A. B. Tutein, and J. A. Harrison, *J. Chem. Phys.* 112, 6472 (2000).
46. D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott, *J. Phys.-Cond. Mat.* 14, 783 (2002).
47. J. Tersoff, *Phys. Rev. B* 37, 6991 (1988).
48. J. Tersoff, *Phys. Rev. Lett.* 61, 2879 (1988).
49. S. M. Foiles, M. I. Baskes, and M. S. Daw, *Phys. Rev. B* 33, 7983 (1986).
50. W. G. Hoover, *Phys. Rev. A* 31, 1695 (1985).
51. G. J. Martyna, M. L. Klein, and M. E. Tuckerman, *J. Chem. Phys.* 97, 2635 (1992).

Received: 19 November 2010. Accepted: 1 December 2010.

Delivered by Ingenta to:  
Tiffany Vargas  
IP : 198.206.219.39  
Sun, 21 Aug 2011 09:41:12

